

Personpersepsjon: Realitet eller sosiale stereotyper?

Eirik Ulveseth



Levert som hovedoppgave ved Psykologisk institutt

UNIVERSITETET I OSLO

20. april 2007

Abstract

Author: Eirik Ulveseth
Title: Personpersepsjon: Realitet eller sosiale stereotyper?
English title: Person perception: Reality or social stereotypes?
Supervisor: Harald Engvik

Consensus refers to the extent to which judges agree in their ratings of a common target. This is a hotly debated area, both empirically and theoretically. The purpose of this study was to examine how consensus would vary as a function of a) how observable traits are, and b) acquaintance. From a critical realist viewpoint, this thesis argues that consensus will assume high values when well operationalized personality inventories are used within a fully crossed design. The data was collected from a leadership program at the Norwegian school of management; this study is a direct result of personality inventories collected there from 1993-2005. The percentage of total variance accounted for by target variance was used as the measure of consensus. Personality judgements from 126 groups, each including 8-11 persons, was analysed. In this sample, the personality inventory showed high levels of structural validity and construct validity. At zero acquaintance, substantial consensus was established for observable traits relating to extraversion. With higher levels of acquaintance, consensus increased substantially for all big five factors, extraversion rising to the level of .65. Consensus seemed to reach a maximum level, seemingly dependent of the quality of the personality relevant information available, and the actuality of different aspects of personality. The findings are discussed within an elaborate evolutionary framework.

Forord

Denne oppgaven er et resultat av et samarbeid med Handelshøyskolen BI i Oslo, Institutt for organisasjon og ledelse. Jeg ble engasjert av BI for å organisere et omfattende datamateriale, en database som jeg også har fått anledning til å studere nærmere.

En rekke personer har vært involvert i meg og mitt liv under denne prosessen, jeg vil rette en spesiell takk til noen av dem.

Det hele hadde ikke vært gjennomførbart uten Harald Engvik. For det første er han ansvarlig for å ha reaktivert min interesse for betydningen av å gjøre gode operasjonaliseringer, og ikke bli forført av ”tallenes magi”. I løpet av de siste to årene har han - i kraft av sitt dyptgripende engasjement - inspirert, kritisert, diskutert og gitt retning til arbeidet. Ettersom fristen har nærmet seg, har det foregått intensiv veiledning per e-post fra utlandet. Takk for ditt vennlige vesen og din suverent pålitelige veiledning.

Jeg ønsker også å takke Paul Moxnes for å ha gjort datamaterialet tilgjengelig for meg. Han har fulgt hele prosessen, og har vært like ivrig på å besvare mine forespørsler som jeg har vært på å avgi dem. Jeg opplever å ha fått innsikt i et kursprogram som jeg selv ikke har deltatt i.

Vemund Vik har utvidet min interesse for samtalens form, pulverkaffe og ”oppgavens struktur”. Takk for lange dager og sene kvelder med og uten innhold, og rikelig form. Lars Ekeland takkes for sitt utsøkte intellekt, som nok en gang har kommet meg til gode. Takk til dere begge for gjennomlesning av manuskriptet.

Til slutt må naturligvis Smukisen takkes for sitt yndige vesen.

Oslo, 20. april 2007

Eirik Ulveseth

Innholdsfortegnelse

1. Innledning.....	5
1.1 Erkjennelsesteoretisk persepsjonsteori.....	6
1.1.1 Konstruktivisme og fenomenologi: Weighted average model (WAM).....	6
1.1.2 Vitenskapelig realisme: Realistic accuracy model (RAM).....	8
1.1.3 Forskjeller mellom modellene.....	10
1.2 Personlighetens eksistens.....	11
1.2.1 Operasjonalisering av personlighet.....	12
1.2.2 Leksikalsk perspektiv på femfaktormodellen.....	13
1.2.3 Utvidet evolusjonistisk perspektiv.....	14
1.2.4 Integrasjon av perspektiver.....	16
1.3 Hypoteser.....	16
2. Metode.....	19
2.1 Oversikt over studiet.....	19
2.2 Subjekter.....	20
2.3 Mål på personlighet og skalaer.....	20
2.4. Prosedyre.....	21
2.5 Valg av analyser.....	22
2.5.1 Faktoranalyse.....	22
2.5.2 Beregning av konsensus.....	23
2.5.3 Anmerkning vedrørende fortolkning av intraklassekorrelasjon.....	25
3. Resultater.....	27
3.1 Faktorstruktur og begrepsvaliditet.....	27
3.2 Observerbarhet og konsensus.....	29
3.3 Konsensus som funksjon av kjennskap.....	29
4. Diskusjon.....	32
4.1 Oppsummering av resultater.....	32
4.2 Faktorstruktur er lik ved begge tidspunkt.....	32
4.3 Konsensus for ukjente er substansiell for trekk som er observerbare.....	33
4.4 Konsensus er høyere for vurdering av kjente.....	35
4.5 Maksimumsnivå for konsensus.....	38
4.6 Styrker og svakheter ved forskningsdesignet.....	40
4.7 Konkluderende betraktninger.....	41
5. Litteraturliste.....	43
6. Appendiks.....	47
6.1 Appendiks A.....	47
6.2 Appendiks B.....	48
6.3 Appendiks C.....	50

1. Innledning

Hvordan vi oppfattes av andre mennesker utgjør en stor del av vår sosiale virkelighet. Ofte kan andres vurdering også ha dramatiske konsekvenser for oss. Den siste tids mediebilde i Norge har blant annet vært preget av saken som omhandler konflikten innad i landsorganisasjonen (LO), og leder Gerd Liv Vallas rolle i denne. Ingunn Yssen var ansatt som internasjonal sekretær, men opplevde samarbeidet med Valla som så vanskelig at hun sa opp sin stilling den 10. januar. Dagen etter offentliggjorde hun oppsigelsesbrevet i dagspressen. Der sto det blant annet:

”LOs leder er både hyggelig, oppmuntrende og generøs med ros så lenge hun blir møtt med hengiven underkastelse. Alle innsigelser, kritikk, selv det som er gitt med de beste intensjoner blir oppfattet som illojalitet. Alt kontrolleres, alle settes opp mot alle og sladring premieres. Det skapes en atmosfære hvor det er lite rom for kreativitet og selvstendige vurderinger.” (Yssen, 2007).

At Ingunn Yssen valgte å sende brevet til media, tyder på en annen agenda: Norges befolkning må få vite hvordan det *egentlig* står til. Og reaksjonene uteble ikke. Fra anonyme kilder dukket det stadig opp informasjon fra den pågående granskningen, og eksperter uttalte seg om Vallas lederstil, hvilke konsekvenser det hele ville komme til å få, og Vallas personlighet. Mitt poeng er at personvurderinger er av avgjørende betydning. Hvilke oppfatninger en person ”rammes av” får konsekvenser, særlig hvis disse oppfatninger sammenfaller med hverandre. Valla var ikke enig i beskrivelsen som ble gitt av henne. Samme dag som Yssen sendte brevet til media, holdt Valla pressekonferanse og hevdet at påstandene ikke hadde ”rot i virkeligheten”. Hvilken virkelighet er det hun snakker om? Det følger av dette at her er det enten en som tar feil og en annen som har rett, eller så henviser de til ulike virkeligheter. Sannsynligvis vil granskningen komme frem til at ”sannheten” ligger et sted midt imellom¹. Dette kan omskrives til et spørsmål om nøyaktighet, samtidig som det poengterer hvor vanskelig det er å oppnå nøyaktige vurderinger. Studiet av personlighetsvurdering, som også er kjent som sosial vurdering, person persepsjon og sosial kognisjon (Funder, 2001), omhandler de vurderinger vi gjør om våre egne og andres

¹ Fra granskningen ble det blant annet konkludert med at ”Vallas forklaring (er) dels preget av begrenset innsikt i, og forståelse av, hvordan hennes atferd påvirker andre, og dels av et forskjønnet bilde av hva som faktisk har skjedd.” (Magnus, 2007). Samme dag som rapporten ble offentliggjort trakk Gerd Liv Valla seg fra sin stilling som LO-leder.

psykologiske egenskaper. Jeg vil i denne studien først og fremst konsentrere meg om bedømmerens virkelighet og det som har fått benevnelsen konsensus (Funder & West, 1993; Shrout, 1993; Kenny, 1994). Konsensus kan defineres som graden av enighet mellom ulike bedømmeres personlighetsbeskrivelser av samme målperson (Target). Er det mulig for en gruppe bedømmere å komme frem til en felles og rimelig sammenfallende beskrivelse av hvordan en person er? Hvilke faktorer kan fremme en slik ”delt” virkelighetsoppfatning? Gitt at dette er mulig: finnes det kriterier for å si at det er en sann – virkelig – beskrivelse? De første spørsmålene skal jeg komme tilbake til når problemstillingene tilspisses. Det siste spørsmålet er i stor grad et spørsmål av ontologisk og erkjennelsesteoretisk karakter. Hvordan er virkeligheten, og hvordan kan vi vite noe om den? Det er mulig å innta en rekke ulike posisjoner til denne problemstillingen.

1.1 Erkjennelsesteoretisk persepsjonsteori

Den vitenskapelige realisme tar utgangspunkt i en erkjennelsesteori som uttrykker at objektene vi persiperer eksisterer, og at de beholder i alle fall noen av sine egenskaper uavhengig av den som observerer. En konstruktivistisk tilnærming står i kontrast til dette, idet den tar utgangspunkt i en fenomenologisk posisjon der verdenen bak sansningen ikke eksisterer, eller i alle fall er det ikke mulig å skille denne fra erfaring og opplevelse (Dancy, 1995; Funder, 1999). Dette er ytterpunkter innen erkjennelsesteori, og det er heller sjelden at en finner disse i ren form innenfor psykologien. Kruglanski (1989) har en god redegjørelse for ulike posisjoner som har vært relevante for studiet av nøyaktighet. Det er hensiktsmessig å ha ytterpunktene med her for å tydeliggjøre hvorfor man innen sosialpsykologien og personlighetspsykologien har studert lignende fenomener på så forskjellige og tilsynelatende uforenelige måter. Jeg vil i det følgende presentere to modeller for nøyaktighet og konsensus innen personpersepsjon med fokus på forskjellene som antydnet her.

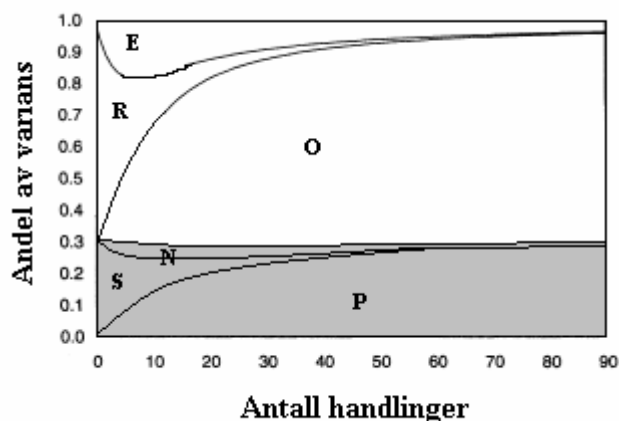
1.1.1 Konstruktivisme og fenomenologi: Weighted average model (WAM)

Den konstruktivistiske, fenomenologiske posisjonen har for en stor grad ligget til grunn for sosialpsykologiens studier av personpersepsjon, mens den vitenskapelige realisme har preget personlighetspsykologiens tilnærming til måling og kartlegging av personlighet (Funder, 1999). Funders bakgrunn for å gjøre en slik inndeling ligger i at den sosialpsykologiske tilnærmingen til person persepsjon har vært konsentrert rundt de kognitive konstruksjoner og sosiale faktorer som påvirker vår persepsjon av personer. Dette har samtidig medført en

nærmest total ignoranse av den sosial realitet (Funder, 1999). Kenny (1994) tilkjennegir også dette, noe han understreker med tittelen på sin bok "Interpersonal perception – a social relations analysis –". Han fremhever at han nettopp ønsker å distansere seg fra den klassiske vignett-forskningen som mislykkes i å fange inn de sentrale elementer i hverdagens personpersepsjon.

"In interpersonal perception research, the perceiver and the target are allowed to interact with each other. Moreover, the two form impressions about each other. In addition to these impressions, self-perceptions and perceptions about how a person thinks others see him or her are often measured. Interpersonal perception research captures the richness of perceiving other people." (Kenny, 1994, s 3).

Kennys organiserer forskningen og sin teoretiske Weighted average model (WAM) omkring ni forskningsspørsmål og en rekke komponenter som sammen, men på ulike måter er ansvarlig for variasjon i konsensus (Kenny, 1994; Funder, 1996). Senere har Kenny forenklet denne modellen til PERSON, et akronym som står for seks kilder til uavhengig varians i personpersepsjon: personality, error, residual, stereotype, opinion og norm (Kenny, 2004). En av Kennys hovedobservasjoner fra forskningen han sammenfatter er at konsensus ikke øker ved økt kjennskap, og han slutter fra dette at "acquaintance is not as important in person perception as generally thought." (Kenny, 2004, s 265). En slik konklusjon gir inntrykk av at våre vurderinger av andre ikke blir mer nøyaktige ettersom vi får mer informasjon over tid. Kennys påstand henger sammen med at relativ target varians (konsensus) består av tre komponenter: Personlighet (P), stereotype (S) og norm (N). Det skraverte feltet i figur 1.1.1 viser hvordan disse forholder seg til hverandre over tid ettersom bedømmer observerer atferd.



Figur 1.1.1 Konsensus utgjøres av det skraverte området, mens R (del av stereotyper som er unik for en bedømmer), O (bedømmers unike syn på target) og E (rest) ikke bidrar til konsensus. P (hvordan en person vanligvis oppfattes), S (delte antakelser basert på informasjon om kategorier) og N (inkonsistens i targets atferd) utgjør konsensus (Tilpasset fra Kenny, 2004, s 271).

Konsensus øker altså ikke over tid, fordi S er høy tidlig i kjennsapsprosessen men avtar ettersom P øker. N (inkonsistens) vil per definisjon ikke korrelere over tid. I denne modellen er det P-komponenten som utgjør et valid kriterium for nøyaktighet, og denne komponenten øker ved økt kjennskap, men oppnår relativt raskt et metningspunkt i Kennys tenkning: “On average, about 15% of the total variance reflects consensus.” (Kenny, 1994, s 204). P er en hypotetisk komponent fordi det aldri er slik at alle personer ser alle handlinger. Hans hovedfokus er på persepsjon, og P vil dermed alltid være en konstruksjon i et ”persiperende felleskap” (Kenny, 2004). Modellen fremstår derfor likevel i nær sammenheng med sosialpsykologiens tradisjonelle tilnærming til personpersepsjon.

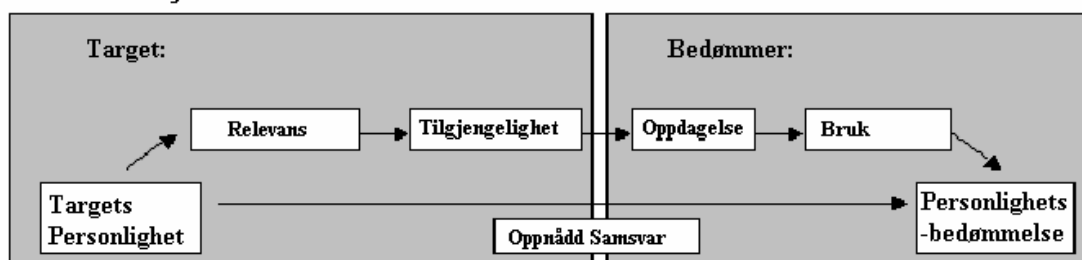
1.1.2 Vitenskapelig realisme: Realistic accuracy model (RAM)

For i det hele tatt å kunne svare på om en vurdering er nøyaktig, trenger vi en definisjon av nøyaktighet som forholder seg direkte til dette spørsmålet (Funder, 1999; 2001). Dette fordrer en realisme som kalles kritisk realisme (Bhaskar, 1989; Cook & Campbell, 1986). Den kritiske realismen har sin basis i en ontologisk teori (Bhaskar, 1989), det begynner altså med et spørsmål om hva som faktisk eksisterer. Funder (1999; 2001) synes å ha som en grunnleggende ontologisk forutsetning at mennesker eksisterer som reelle enheter i en virkelig, sosial verden hvor de påvirker og blir påvirket. Fra denne forutsetningen følger

spørsmålet om hvordan vi kan få kunnskap om dette som eksisterer, og den kritiske realismen impliserer at erkjennelsen (her: nøyaktighet) er å finne i skjæringspunktet mellom ulike operasjonaliseringer av virkeligheten (Funder, 1999; Kruglanski, 1989). En kan si at en vurdering er nøyaktig i den grad den kan predikere atferd, viser konsensus (mellom observatørvurderinger) og enighet (mellom egenvurdering og observatørvurdering). Dette kan omformuleres til et spørsmål om begrepsvalidering: Personlighetsvurderinger må valideres gjennom prediktiv validering (Cook & Campbell, 1986; Kenny, Horner, Kashy & Chu, 1992), konvergent og divergent validering (Campbell & Fiske, 1959; Engvik, 1993).

Funders ”Realistic accuracy modell” (Funder, 1999; 2001) bygger på tre grunnleggende forutsetninger. Det antas at personlighetstrekk eksisterer, at personer vurderer personlighetstrekk (i det minste noen ganger), og at slike vurderinger kan være nøyaktige (i det minste noen ganger). Dersom en kan gå med på disse grunnantagelsene, åpnes det opp for spørsmålene: hvordan kan en bedømmelse være nøyaktig, og hvilke faktorer kan fremme en nøyaktig bedømmelse av personlighetstrekk? Funders modell uttrykker at nøyaktig bedømmelse av trekk er vanskelig fordi *relevant* atferd må være *tilgjengelig* og bedømmeren må *oppdage* og *bruke* holdepunktene i sin vurdering (Funder, 1999; 2001). Figur 1.1.2 gir en illustrasjon av modellen.

Realistic Accuracy Model:



Figur 1.1. 2. Realistic Accuracy Model viser prosessen mot en nøyaktig bedømmelse. For det første må targets personlighet avgi relevante eller observerbare atferdskorrelater. Neste trinn deler inn relevante atferder i hvor sansbare eller tilgjengelige de er hos target. Tredje steg i prosessen handler om bedømmerens evne til å legge merke til den relevante og tilgjengelige atferd, og det siste leddet omhandler i hvilken grad bedømmer er i stand til å benytte denne informasjonen korrekt, det vil si å knytte den til personlighetstrekket den representerer (Tilpasset fra Funder, 2001, s136).

Dette kan også representeres i formell form: Nøyaktig vurdering = [(tegn fra atferd som er relevant for et personlighetstrekk) X (i hvilken grad disse tegn er tilgjengelige for observasjon)] X [(i hvilken grad disse tegn oppdages) X (på hvilken måte tegne brukes)] (Funder, 1999). Funder har kvantifisert disse med maksimumsverdier på 1 og minimumsverdier på 0 for å understreke at nøyaktighet er vanskelig. Dersom prosessen svikter på et punkt og antar verdien 0, vil også nøyaktigheten bli lik 0. Slik sett er realistisk nøyaktighet først og fremst et filosofisk ideal. For å nærme oss dette idealet må vi være opptatt av alle faktorene i modellen. Funder opererer med fire moderatorvariabler som korresponderer til modellen: 1. Good Trait: trekkets observerbarhet. 2. Good Target: targets observerbarhet. 3. Good information: Kvalitet og kvantitet av informasjon som oppdages (kjennskap) og 4. Good Jugde: bedømmerevne (Funder, 1999). Funders modell organiserer mye forskning på personpersepsjon og beslutningsprosesser.

1.1.3 Forskjeller mellom modellene

Grunnforskjellen mellom de to modellene jeg har presentert ligger i de avhengige variablene de er konstruert for å forklare (Funder, 1996). I WAM er den avhengige variabel *konsensus*. I RAM er den avhengige variabel *nøyaktighet*. En del av konsensusen oppstår på grunn av gjensidig nøyaktighet innenfor Kennys modell (1994), mens i Funders modell er konsensus en feilbarlig kilde som *kan bety* at en oppfatning er nøyaktig (Funder, 1996). Kenny og Funder har til dels motstridende fokus: "It seems unlikely that all the perceivers could be mistaken all of the time, *however it is possible*." (Kenny, 2004, s 269, min uthevelse). Kennys vektlegging fastholder en forbindelse til sosialpsykologiens konstruktivistiske tradisjon der interessen for persepsjon i seg selv har forrang fremfor virkelige forhold (Kenny, 2004; Funder, 1996; 1999). Det er naturligvis slik at til og med våre mest fundamentale orienterende oppfatninger kan diskuteres og dermed gir oss et usikkert utgangspunkt (Funder, 1999), men et slikt fokus mislykkes i å anerkjenne at en denne usikkerheten ikke bare rammer studiet av nøyaktige personvurderinger: "its a basic fact of life" (Funder, 1999, s 9). Selve *begrepet nøyaktighet fordrer realisme*, i alle fall synes det rimelig å hevde at det er mer hensiktsmessig. Hypotesene i dette studiet er utledet fra Funders modell, før jeg går inn på det er det nødvendig å si noe om hans grunnantagelse.

1.2 Personlighetens eksistens

”I conclude from these three inquiries that the character which shapes our conduct is a definite and durable “something,” and therefore that it is reasonable to attempt to measure it.” (Galton, 1884, s 181).

Dette tilsynelatende rimelige utgangspunktet som i sin tid inspirerte studiet av individuelle forskjeller, har vært gjenstand for en lang kontrovers som strakk seg over mer enn to tiår (for oversikter se Funder, 1999; Kendrick & Funder, 1988; Epstein & O’Brien, 1985). En av hovedaktørene, Walter Mischel (1968), oppsummerte en rekke studier som viste tilsynelatende lave korrelasjoner mellom personlighetstrekk og atferd, og tidsånden var tilrettelagt for å ta imot hans hovedkonklusjon: personlighetstrekk er av ekstremt begrenset nytte for prediksjon av atferd, hva en person gjør avhenger mest av de eksakte omgivelser og faktorer i situasjonen (Mischel, 1968; Epstein & O’Brien, 1985).

Personlighetspsykologien, og især nøyaktighetsforskningen, var på dette tidspunkt inne i en dyp krise (Funder, 1999). Dette skyldtes særlig Cronbachs (1955) kritikk av metodene man brukte. Egenvurdering ble regnet som kriterium for nøyaktighet, og andres vurderinger ble ansett som nøyaktige i den grad de korresponderte med dem. Cronbachs kritikk omhandlet dette kriteriet, man hadde ikke tatt høyde for at profilkorrelasjoner mellom egenvurdering og bedømmervurdering ikke bare er forårsaket av at vurderingen er nøyaktig. En vurdering består av flere komponenter, og særlig to av disse komponentene representerer ikke reell nøyaktighet i en bedømmelse (Cronbach, 1955). Han viste at korrelasjon kan oppstå på bakgrunn av indre forestillinger som target og bedømmer deler, som for eksempel hvilke trekk som er sosialt ønskværdige. Disse forestillingene kan korrespondere, og da står vi i fare for å få en korrelasjon mellom bedømmelser som ikke knytter seg til faktisk observasjon av target. Han foreslo at denne komponenten kunne kalles ”accuracy in predicting the generalized other” (Cronbach, 1955, s 179). Han viste også at korrelasjon kan oppstå på bakgrunn av delt responsstil mellom to bedømmere. Dette representerer heller ikke nøyaktighet, fordi det knytter seg til en måte å svare på, og ikke til faktisk observasjon av target. Denne komponenten refereres til som ”elevation” (Cronbach, 1955).

I dag er disse vanskene i stor grad overvunnet ved hjelp av både mer sofistikerte psykometriske prosedyrer (Shrout & Fleiss, 1979; Shrout, 1993; Kenny, 1994), bedre forståelse av hva effektstørrelser faktisk representerer (Rosenthal & Rubin, 1982) og forskningsdesign som er bedre tilrettelagt etter problemstillingene man ønsker å besvare (Kenny, 1994; 2004; Funder, 1999; 2001; Kenny et al., 1992). Men på femtitallet hadde man ikke disse metodiske nyansene, og ble stående igjen med et inntrykk av at de nøyaktighetsestimaterne som forskningen brakte frem var feilaktige. Konsekvensen var at studiet av nøyaktighet ble forlatt (Funder, 1999; 2001), eller i alle fall ble mengden studier omkring temaet kraftig redusert (Kruglanski, 1989). Først etter 20 år kom det en artikkel som kunne argumentere for at den langvarige person-situasjon kontroversen ikke lenger var gyldig (Kendrick & Funder, 1988). Senere har en rekke aktører innrømmet at kontroversen har bygget på en falsk dikotomi mellom personlighetsmessige og situasjonelle determinanter for atferd (Funder, 1999; 2001; 2006; Kenny, 1994; 2004; Mischel & Shoda, 1995; Kammrath, Mendoza-Denton & Mischel, 2005; Epstein & O'Brien, 1985). I dag er det bred enighet om at atferd er en funksjon av interaksjonen mellom person og situasjon: "Nowadays, everybody is an interactionist" (Funder, 2006, s 22).

1.2.1 Operasjonalisering av personlighet

Selv om studiet av personlighet ikke lenger er definert ut av den vitenskaplige psykologien, er det likevel en utfordring å operasjonalisere personlighet på valide måter. Dette elementet er også avgjørende for nøyaktighet. Dersom vi kunne gått inn og direkte målt persepsjonen av andre mennesker på en fullstendig fenomennær måte (slik blant annet Wilhelm Wundt forsøkte på 1880-tallet; se Saugstad, 1998) kan det hende at to personers oppfatninger ville sammenfalle selv om de benyttet divergerende beskrivelser av sin opplevelse. Det finnes ikke noe enkelt valid kriterium for opplevelse, eller for personlighet. Campbell og Fiske (1959) fremhever dette og foreslår en praktisk metodologi for forskningen. Den beste måten å nærme seg fenomener på, er å måle operasjonaliseringene (for eksempel trekk) med flere metoder² og se dem sammen i det de har kalt en multitrekk-multimetode matrise. Fra korrelasjonene i matrisen kan man gjøre slutninger om begrepsvaliditet: Operasjonaliseringer av personlighet er valide i den grad teoretisk beslektede trekk (for eksempel pratsom og aktiv) også henger sammen empirisk (konvergent validitet), og at trekk vi hevder er teoretisk uavhengige viser

² Når mange bedømmere vurderer det samme target med mange trekk, kan hver bedømmer betraktes som en metode. Dette er utgangspunktet for konsensus, slik har konsensus en direkte forbindelse til reliabilitet.

lav eller fraværende korrelasjon (divergent validitet). Med utgangspunkt i den menneskelige bedømmer kan dette omsettes til et krav om intersubjektivitet. Mener vi det samme med det vi sier?

1.2.2 Leksikalsk perspektiv på femfaktormodellen

Den leksikalske tradisjonen har et intersubjektivt utgangspunkt i språket. Mange har fulgt i Galtons (1884) fotspor og gjort dypdykk inn i språket ut fra en tankegang om at alle tydelige og sosialt relevante individuelle forskjeller mellom personer over tid vil bli kodet inn i det språket de bruker. Dette har blitt kjent som den leksikale hypotese, og utgjør rammen for forskningstradisjonen som befatter seg med studiet av individuelle forskjeller (Goldberg, 1981; 1993; Saucier & Goldberg, 1996; John & Shrivastava, 1999). Logikken følges videre i identifikasjonen av viktige trekk-beskrivende ord: jo viktigere ett personlighetsattributt er, desto flere synonymer og distinkte fasetter ved attributtet vil finnes innen et språk, og de viktigste individuelle forskjeller vil ha korresponderende ord i alle språk (Saucier & Goldberg, 1996). Goldberg (1993) har gitt en grundig fremstilling av de historiske aspektene ved utviklingen av femfaktormodellen, hvem som var først og hvor de gikk feil, kontroverser om antall faktorer, navn på faktorene, og ikke minst; hva betyr de egentlig? Her kommer et sentralt og viktig skille. Nemlig om de fem faktorer skal forstås som en hierarkisk (og aggregert) taksonomi av personlighetsattributter som sådan har implikasjoner for studiet av individuelle forskjeller (for eksempel Norman, 1963; Goldberg, 1993), eller om de fem faktorene er fenotypiske uttrykk for genotypiske trekk (for eksempel McCrae & Costa, 1999). Saucier og Goldberg (1996) understreker at dette er to distinkte vitenskaplige spørsmål, som rimeligvis fordrer forskjellige forskningsprosedyrer.

En av de mest vedvarende kritikkene som har vært rettet mot bruk av personlighetstrekk er at trekk ikke gir noen forklaring (Funder, 1999). Det er rimelig å tenke seg personlighetstrekk som merkelapper for konsistente mønstre av tanker, følelser og handlinger (Hogan, 1996), men trekk gjengir fenomenet uten å bringe oss nærmere en forståelse av det. Denne kritikken har utgangspunkt i den filosofiske distinksjonen mellom vertikale og horisontale forklaringer (Funder, 1999). Horisontale forklaringer handler om i hvilken grad et konsept forener minst to manifeste fenomener, kanskje især fenomener som har avstand fra hverandre (Funder, 1999). Hydrogen og oksygen har svært forskjellig "fenomenologi". De forenes under "vann", et livsnødvendig konsept som blir forstått av alle. Dersom vi finner at personer som går på

mange fester også er i besittelse av en hel masse positive emosjoner, så gir ideen om ekstraversjon en horisontal forklaring, idet man hevder at begge deler er manifestasjoner av det samme trekket. Sier dette oss så noe mer? Det kommer an på hvilket spørsmål man er interessert i å besvare. En slik måte å benytte trekkbeskrivelser på er relevant for menneskers mest opplagte og vanlige formål med å benytte trekkbeskrivelser: "what to do next?" (Funder, 1999, s 53). Følgende illustrasjon kan tjene som eksempel (den bygger på Funder, 1999, s 53-54).

Jeg opplever at en person behandler meg foraktfult. Er han bare foraktfull mot meg, eller er han foraktfull mot meg *på grunn av* noe jeg har gjort? Dette er et viktig spørsmål, og svaret på det er et svært nyttig forklarende trinn. For å besvare det behøver jeg mer horisontal informasjon: Er denne personen foraktfull mot alle andre? Eller bare meg? Behandler alle meg slik, eller er det bare ham? Når jeg har besvart disse spørsmålene, kan jeg avgjøre om hans atferd kommer fra hans personlighet eller ikke. Jeg undersøker litt, og finner at de aller fleste oppfatter ham som foraktfull, han behandler alle veldig dårlig. Uavhengig av om dette er "slik han virkelig er" (et interessant spørsmål i og for seg) har jeg nå fått en forklaring til det folk vanligvis trenger forklaringer for: Hva gjør jeg nå? Jeg holder meg langt unna ham! Hans atferd kan forklares ut ifra en rekke "dypere" vertikale forklaringsmodeller, men det første trinnet er alltid en horisontal konklusjon som i dette tilfellet er at det er noe ved ham og hans væremåter (beskrevet med ordet "foraktfull") som er kilden til hans atferd mot meg. Derfor kan man ikke undervurdere viktigheten av den horisontale forklaring. Sammenbindingen av en atferdsobservasjon med en annen gjennom å tilskrive et felles trekk til dem begge (Funder, 1999), har et implisitt krav om nøyaktighet fordi hele prosessen forutsetter at jeg kan stole på min vurdering.

1.2.3 Utvidet evolusjonistisk perspektiv

Det finnes en rekke relevante vertikale forklaringsmodeller for personlighet. Både Buss (1996), Hogan (1996) og Wiggins og Trapnell (1996) har foreslått evolusjonpsykologiske perspektiver som er velegnet for å gi femfaktormodellen et "dypere" innhold. Hogans sosioanalytiske tilnærming er ytterligere relevant for konsensus, fordi han spesifikt presenterer en forståelse av femfaktormodellen der observatørens perspektiv står i fokus. Samtidig er forbindelsen sterk til den leksikalske tradisjonen som også har utgangspunkt i tanken om evolusjon (Goldberg, 1981; 1993; Saucier & Goldberg, 1996). Som mennesker ble

vi utviklet innen grupper, og vi er kulturanvendere. Enhver gruppe har et statushierarki, og hierarkiets regler er innarbeidet i kulturen. Vår evolusjonære historie foreslår at mennesker trenger *sosial status* og *sosial aksept*. Reproductiv suksess er bunnlinjen i evolusjonsteori (Darwin, 1859). Jo mer status og aksept en person nyter, desto bedre vil han/hun gjøre det "in the game of life" (Hogan, 1996, s 174). I denne forstand er mennesket tvunget til sosial interaksjon, fordi status og aksept er noe vi får *fra andre*. Den sentrale hypotesen som knytter seg til femfaktormodellen, er at den inneholder de kategorier som mennesker benytter for å vurdere andre personers grad av sosial status og aksept. Wiggins og Trapnell (1996) gir en prioritet til de to første faktorene (extraversion og agreeableness), en prioritering som stammer fra deres overbevisning om at disse statistiske abstraksjonene korresponderer nærmest med deres begrepsdistinksjon mellom "communion" og "agency" (Wiggins & Trapnell, 1996). Denne distinksjonen tar mange former (jamfør sosial status og sosial aksept), noe som gjenspeiler at det knytter seg til noe helt fundamentalt i den menneskelige eksistens (Wiggins & Trapnell, 1996). Gjennom en andreordens faktoranalyse sannsynliggjør Digman (1997) dette. I analysen inngikk 14 studier med robuste femfaktorstrukturer. Hans ønske var å skape et bredt rammeverk for "the myriad (of) personality constructs that have been offered by theorists..." (Digman, 1997, s 1246). Han finner to hovedfaktorer, kalt alpha og Beta. I alle studiene hadde faktor II, III, og IV høyeste ladninger på alpha, mens faktor I og faktor V hadde høyest ladninger på Beta. "Communion" og "sosial aksept" relaterer seg til alfa, mens "agency" og "sosial status" relaterer seg til Beta (Digman, 1997; se også Wiggins & Trapnell, 1996).

Buss (1996) utvider perspektivet ytterligere ved å understreke den gjennomgripende funksjonalismen som ligger innebygd i et evolusjonsperspektiv. Å identifisere en mekanismes funksjon er essensielt for å forstå dens natur: hvordan den fungerer, hvilke kontekster den aktiveres i, og hvorfor den eksisterer i det hele tatt. Mennesket er en kompleks sammensetning av integrerte mekanismer som har bestått gjennom naturlig seleksjon (Darwin, 1859; Buss, 1996). Mekanismene har bestått fordi de er adaptive: de løser problemene. Reproductivt viktige ressurser som mat, territorium og attråverdige partnerer vil typisk tilfalle dem som er høyt i det sosiale hierarkiet. Siden disse adaptive problemer har et innebygd sosialt aspekt, vil også våre psykologiske mekanismer være utformet for å løse sosiale problemer. Evolusjon opererer relativt i den forstand at viktige adaptive ressurser er begrenset sammenlignet med antall personer som ønsker dem (Darwin, 1859). En persons

oppnåelse vil nødvendigvis gå på bekostning av andre, vi er konkurrenter (og allierte) innen våre respektive ”sosialadaptive landskap” (Buss, 1996).

1.2.4 Integrasjon av perspektiver

Buss (1996) finner bred støtte for at våre mellommenneskelige strategier varierer og lar seg predikere som funksjon av spesifikke sosialadaptive problemer, og Hogan (1996) fremhever at sosial status og aksept er gjennomgripende og særlig viktige for sosial adaptasjon. Wiggins og Trapnell (1996) gir faktor I og faktor II forrang fordi de korresponderer nærmest med disse overordnede evolusjonistiske begrepene, og Digman (1997) utvider dette igjen ved å vise statistisk at de andre faktorene også knytter seg til disse overordnede begrepene. Dette får noen følger for observerbarhet. Hvilke trekk som til enhver tid er observerbare hos andre, vil være avhengig av hvilke sosialadaptive problemer som er aktualisert. Ifølge Funders modell fremmes nøyaktige personvurderinger av at trekk er observerbare. Fra perspektivet redegjort for her er det en nær sammenheng mellom observerbarhet og mekanismene vi har utviklet for å løse sosialadaptive problemer. Her blir individuelle forskjeller av avgjørende betydning, fordi mekanismene nettopp har som funksjon å oppdage potensielle konkurrenter (status) og potensielle allierte (aksept). Trekkbeskrivelsene i dette studiet er forsøkt optimalisert i forhold til disse aspektene.

Faktorene i den leksikalske tradisjonen benevnes med romertall, og rekkefølgen gjenspeiler hvor mye variasjon hver faktor forklarer i leksikalske undersøkelser av personlighetsbeskrivende ord (Saucier & Goldberg, 1996). En fordel med å benytte romertall, er at faktorene er sammensatt av flere egenskaper, og det er derfor vanskelig å finne dekkende betegnelser (Engvik & Føllesdal, 2005). Samtidig finnes det noen godt innarbeidede betegnelser i den engelskspråklige forskningslitteraturen, jeg vil i fortsettelsen benytte følgende oversettelser: I. Ekstraversjon, II. Varme³, III. Samvittighetsfullhet, IV. Emosjonell stabilitet (versus nevrotisisme), V. Intellekt. (Etter Engvik & Føllesdal, 2005).

1.3 Hypoteser

Som indikert i innledningen har RAM blitt benyttet for å utlede forskningshypotesene i dette studiet. Operasjonaliseringen av personlighet har blitt gjort med samme instrument ved to tidspunkt. Jeg vil i fortsettelsen referere til de ulike *tidspunktene* med ”vurdering av ukjente

³ Det finnes ikke noen god norsk oversettelse på ”agreeableness”. Jeg har valgt å benytte ”varme”, betegnelsen som opprinnelig ble benyttet for faktor II i 5PFa (Engvik, 1993).

personer” og ”vurdering av kjente”⁴. Dersom spørreskjemaet gir forventet femfaktorstruktur ved begge tidspunkt, gir det et godt utgangspunkt i å forklare forskjeller mellom tidspunktene. Det forutsettes at ordene (trekkene) betyr det samme, altså beholder de sin intersubjektive mening. Forskjeller i konsensus over trekk og faktorer har dermed ikke sitt utgangspunkt i uenighet om språket vi bruker for å beskrive, men i hvordan targets skal beskrives med språket. Det forventes at selve språket vil være stabilt. Altså fem faktorer med tilfredsstillende konvergent og divergent validitet som forklarer en rimelig andel av variansen i utvalget. Første hypotese foreslår dette:

H1: Faktorstruktur er lik ved begge tidspunkt.

Ifølge RAM vil et godt trekk være et trekk som er relevant og tilgjengelig. Trekk som ”pratsom” er lett å bedømme fordi direkte relevant atferd er – per definisjon – ofte tilgjengelig for andre. I kontrast står trekk som ”bekymret” som man må slutte seg til fra for eksempel ansiktsuttrykk, gester og sukk – ting som også kan tolkes på en rekke måter -. Effekten av observerbarhet kan synes åpenbar, men en kontrasterende hypotese er at konsensus er høy ved lavt kjennskap på grunn av sosialt delte stereotypier (Kennys S). Dersom det er riktig, er det ingen grunn til å anta at observerbare trekk skulle ha høyere konsensus – det ”persiperende felleskap” kunne like gjerne konstruere rykter om bekymring som om snakksalighet. Et ytterligere aspekt fra det evolusjonistiske perspektivet jeg har redegjort for, er at et trekkets relevans også avgjøres av hvilke aspekter i situasjonen de ulike personer opplever som betydningsfulle for å løse aktuelle sosialadaptive problemer. For vurdering av ukjente vil følgelig prosedyren forut for personvurderingene ha relativt større innvirkning enn ved vurdering av kjente. Dette vil undersøkes med utgangspunkt i følgende hypotese:

H2: konsensus for ukjente er substansiell for trekk som er observerbare.

Ifølge Kenny (1994, 2004) leder ikke mer kjennskap til høyere konsensus. Dersom det er riktig, har det dramatiske konsekvenser for en rekke områder. En klinisk psykologs evne til å vurdere endring i personlighetsstruktur som et resultat av terapi, og tanken om at en kan vite mye om en nær venns personlighet, vil begge deler være behagelige illusjoner som ikke bekreftes empirisk. Funder (1999; 2001) opererer med variabelen ”Good Jugde”, som i RAM

⁴ Denne refereringen er i en forstand problematisk, siden hypotesene undersøker disse aspektene. Jeg finner det likevel hensiktsmessig å bruke en enkel form.

er den bedømmer som har kapasitet til å avdekke og benytte tilgjengelige trekk på korrekte måter. Ved siden av de metodiske vanskene, kan selve viktigheten av nøyaktig personpersepsjon være med på å forklare hvorfor det har vist seg overraskende vanskelig å finne forskjeller mellom bedømmeres evne til å vurdere personlighet (Funder, 1999). De grunnleggende oppgavene i personpersepsjon er *så viktige at nesten alle klarer det*. Denne siden av modellen undersøkes ikke i dette studiet, men det forutsettes at utvalget er en relativt homogen gruppe ”good jugdes”. Det følger da at eventuell variasjon i konsensus over tid knytter seg til kjennskap, selv om det ikke kan utelukkes at de over tid (for eksempel gjennom kursprogrammet) har blitt bedre ”menneskekjennere”⁵.

I forlengelsen av dette er det interessant å se nærmere på *hvordan* kjennskap virker inn i vurdering av andres personlighet. Funder fremhever to aspekter ved ”good information”: kvantitet og kvalitet. Jeg skal komme tilbake til disse aspektene i diskusjon av resultater. Ut fra RAM vil observerbare trekk, ettersom man blir bedre kjent, få mindre betydning. Det innebærer ikke at effekten av observerbarhet (relevans og tilgjengelighet) blir borte, det betyr at ettersom kjennskap øker vil flere og mer subtile trekk som har sosialadaptive funksjoner bli tydelige for bedømmere. Fra disse betraktningene følger hypotesen:

H3: Konsensus er høyere for kjente enn for ukjente.

En siste relevant diskusjon knytter seg til hvor raskt vi kommer frem til høyest mulig nivå av konsensus. Tradisjonelt har man sett personpersepsjon som en relativt saktegående og gradvis prosess. Sosial penetreringsteori (Altman og Taylor, 1987) benytter et velkjent bilde for å beskrive prosessen. Å bli kjent med en person kan lignes med det å skrelle en løk: Ettersom en person viser seg selv ”lag for lag” over tid, blir det mulig for andre å bli kjent med og komme nærmere personens innerste kjerne. I motsetning til dette fremhever Kenny at nøyaktig kjennskap (P-faktoren) har oppnådd sin maksimumseffekt på konsensus etter noen få timers observasjon (Kenny, 2004). Han benytter samme metafor, men vektlegger et annet aspekt ved løken: ”The distinctive taste of an onion is as marked in its outer layer as it is in the innermost layer”. (Kenny, 2004, s 277). Den kvantifiserte versjonen av RAM indikerer at nøyaktighet øker som funksjon av at relevant og tilgjengelig atferd blir oppfattet og brukt

⁵ Dette kan se ut som en tautologi, men er relevant for generalisering. Ifølge RAM kan økt nøyaktighet være konsekvens både av kvaliteter ved target og bedømmer. Hvis det er slik at man, ved å ta et kurs på BI, blir ”good jugde”, så er det problematisk og generaliserer til populasjoner som ikke har fått slik ”treatment”.

riktig. Dersom ”Good information” (kjennskap) er en viktig moderator i denne prosessen, er det rimelig å forvente følgende tidstrend i materialet:

H4: Konsensus fortsetter å øke ettersom mer informasjon om targets blir tilgjengelig.

2. Metode

2.1 Oversikt over studiet

Denne forskningen har innebåret omfattende kvantitativ personlighetskartlegging som kan karakteriseres i følgende punkter:

- a) Utvalget er stort, og data har blitt samlet inn i en årrekke fra 1994-2005. Dette gir gode forutsetninger for å anta en sterk ytre validitet, noe som gir en generell stabilitet til empiriske funn. Innsamlingen av personlighetsvurderinger var organisert i grupper på 8-11 personer der gruppemedlemmene både er targets og bedømmer i et fullstendig krysset design (Round-robin design, Kenny, 1994). Et forsiktig estimat gir omtrent 12000 personlighetsbedømmelser ved hvert tidspunkt, altså 24000 personlighetsbedømmelser.
- b) Hver år har det blitt benyttet det samme instrument for personlighetsvurdering. Det er en kortversjon av en adjektivbasert kollegavurderingsinstrument som tidligere har vist seg og korrelere høyt med andre mål på femfaktorbasert personlighet (Engvik, 1993). Kortversjonen er valgt ut på bakgrunn av tidligere undersøkelser, og består av de adjektivene som har vist høy intersubjektivitet samtidig med at de gir høy strukturell validitet.
- c) Det har blitt forsikret at personer ikke kjenner hverandre fra før. Det ble gitt muntlig instruks om at man ikke skal vurdere personer man har kjennskap til fra før, og dette stod skriftlig øverst på hvert spørreskjema. I tillegg ble skjemaene i etterkant kodet i kategorier slik at de vurderingene som likevel har blitt gjort av kjente personer ble kodet som sådan.
- d) I løpet av årene har det blitt foretatt kontrollerte variasjoner i grad av kjennskap. Flest grupper har – gitt prosedyren for innsamling – ingen tidligere kjennskap til hverandre. I noen grupper har alle personer fått anledning til å presentere seg i ti sekunder for resten av gruppen. Et år ble første personvurdering samlet etter 4 fulle skoledager. Vurderinger på slutten av kurset har alltid blitt gjort på siste dag.

2.2 Subjekter

Utvalget er hentet fra et lederutdanningsprogram ved handelshøyskolen BI, institutt for organisasjon og ledelse, ved skolene i Oslo, Bergen, Stavanger og Trondheim. Data er samlet inn i årene 1994-2005. Programmet gjennomføres over to semestre, hele perioden strekker seg over 7 måneder med fem samlinger à tre-fem dager. Sentrale opptakskrav er minimum tre års høyere utdanning fra universitet eller høyskole, minimum fire års yrkeserfaring, og alder fra 25 år og oppover (ingen øvre grense). Lang og relevant arbeidserfaring kan gi dispensasjon fra kravet om offentlig godkjente studiepoeng. Et utdrag fra kursbeskrivelsen – som bare har vært utsatt for mindre revisjoner gjennom årene - sier noe om hva kurset handler om og dermed også en indirekte beskrivelse av utvalget.

”Det overordnede mål for programmet er å formidle en helhetlig forståelse av hvilke krefter i det menneskelige samspillet som ligger til grunn for trivsel, effektivitet og fornying. Gjennom en kontinuerlig prosess formidler programmet en fagpsykologisk viten som gir den enkelte deltaker en unik mulighet også for personlig utvikling.”
(Samspill og ledelse; anvendt organisasjonspsykologi, 2007).

I en kartlegging av effekten av studieprogrammet fremkommer det hvilken type stilling studentene hadde da de tok ”samspill og ledelse”. Omtrent 30 prosent var ikke ledere, 33 prosent var mellomledere, 25 prosent var prosjektleder eller personalleder, mens de resterende 12 prosent befant seg i topplederposisjoner. Altså 70 prosent ledere og 30 prosent ikke-ledere. Kjønnfordelingen har variert noe fra år til år og fra sted til sted, men i gjennomsnitt har fordelingen vært 50/50. Kurset defineres som et lederprogram med undertittel ”anvendt organisasjonspsykologi”, noe som gjør det rimelig å anta at de som ikke var ledere ved kursets start har ambisjoner om dette. Samtlige studenter var norskspråklige.

2.3 Mål på personlighet og skalaer

Kollegavurdering har i alle årene blitt foretatt med 20 adjektivskalaer plukket ut fra 5PFa (fem personlighetsfaktorer, adjektivversjon; Engvik, 1993). 5PFa har vist seg å være en god operasjonalisering av fem store faktorer som korresponderer til femfaktormodellen (Engvik, 1993). Bakgrunnen for å benytte en kortversjon har vært at det er mulig å vurdere mange personer på relativt kort tid. Samtidig tilsvarer antallet på 20 ledd formatet til Norman (1963), et av de tidlige studiene der det ble funnet femfaktorstruktur for vurdering av ukjente

personer. En rekke studier har benyttet lignende operasjonaliseringer av femfaktormodellen, forutsetningen for å sammenligne resultatene er altså gode. Leddene i kortversjonen (heretter 5PF20a) ble valgt ut med bakgrunn i intersubjektivitet, det vil si de leddene som gav høyest konvergent og divergent validitet i fire undersøkelser der 5PFa har blitt brukt (Engvik, 1993). Samtidig ble det etterstrebet at leddene gav høy strukturell validitet for fem faktorer⁶ (Engvik, 2007, personlig kommunikasjon). Faktorer med tilhørende adjektiver er gjengitt nedenfor.

- I. Ekstraversjon: (1 aktiv, 2 pratsom, 3 dominerende, 4 utadvendt).
- II. Varme: (5 varm, 6 ikke kynisk, 7 omsorgsfull, 8 vennlig).
- III. Samvittighetsfullhet: (9 velorganisert, 10 strukturert, 11 planmessig, 12 disiplinert).
- IV. Emosjonell stabilitet: (13 ubekymret, 14 ikke hårsår, 15 ikke nervøs, 16 ikke anspent).
- V. Intellett: (17 original, 18 kreativ, 19 fantasifull, 20 oppfinnsom).

Adjektivene ble presentert på 7-punkts skala i balansert rekkefølge, og adjektivenes antatt positive pol ble rotert. Begge deler for å minimere effekten av responsstil. To uavhengige kommisjoner, hver på 7 profesjonsstudenter i psykologi, rangerte sammen med førsteamanuensis Harald Engvik adjektivene etter hvor observerbare de er. Det var ønskelig med en skala på dette fordi observerbarhet relaterer seg direkte til RAM og flere av hypotesene.

2.4. Prosedyre

Måten data har blitt samlet inn på har vært den samme fra år til år, med noen variasjoner. Prosedyrene for innsamling av førsteinntrykk og vurdering av kjente var forskjellige. Gruppeinndelingen som gir rammen for førsteinntrykksvurderinger vedvarer hele kurstiden, altså en strukturell grunn til å anta at personer innen grupper over tid vil ha høyere kjennskap enn personer over grupper.

Kurset starter med at lærerne har en åpning og introduksjon til kurset, omtrent ti minutters varighet. Underveis henges det opp en fargeplansje på tavlen med fem farger (rød, sort/hvit,

⁶ Dette innebærer at fem faktorer forklarer en rimelig mengde variasjon i datamaterialet. Det viste seg at intersubjektivitetskravet og kravet om strukturell validitet i stor grad overlappet. De ledd som gav best konvergens og divergens, var også de som aggregert (faktorer) forklarte mest varians.

gul, grønn og blå), og etter introduksjonen er ferdig skal kursdeltakerne velge en farge de foretrekker. Når alle har valgt, skal de rangere fargene og begrunne sitt første og sistevalg skriftlig med 3-5 stikkord. De får så beskjed om å – uten å snakke sammen - fordele seg rundt i rommet rundt plakater som har fargen med deres foretrukne førstevalg. Når det er like mange personer i hver gruppe organiseres de i en sirkel og skal vurdere alle de andre personene i sin gruppe med adjektivskjema. De instrueres om å ikke fylle ut skjema for personer de kjenner fra før. Først når dette er ferdig får de anledning til å kommunisere verbalt med hverandre. Hele prosedyren tar omtrent 50 minutter.

Stemningen i rommet er varierende, under utfylling følges alltid instruksjonen om å ikke kommunisere verbalt. Under valg av farger og gruppefordeling blir det noe høyere lydnivå fra studentene; noen ler, andre blir alvorlige, noen kommenterer, og lærerne kan påpeke når det fortsatt er for mange personer i en gruppe. Typisk fordeler studentene seg slik at rød, blå, og grønn gruppe har for mange personer, mens gul og hvit/sort gruppe har for få eller ingen personer.

I 1994 fikk studentene anledning til å presentere seg i 10 sekunder for gruppen før utfylling av spørreskjema. I 1999 ble ”førsteintrykket” samlet på siste dag på første modul, altså etter at studentene hadde tilbrakt fire fulle studiedager sammen. Konsensusestimaterne for disse gruppene vil følgelig holdes for seg selv.

Siste kursdag på siste modul (etter syv måneder) samles gruppene igjen og fyller ut adjektivskjema for de samme personene som ved kursets start. Også nå understrekes det at de ikke skal fylle ut for personer de kjente fra før.

2.5 Valg av analyser

2.5.1 Faktoranalyse

En forutsetning for å kunne sammenligne resultater på faktornivå ved to forskjellige tidspunkt, er at begge datasettene viser forventet femfaktorstruktur. Alle bedømmelser med adjektivskalaene ble for hvert tidspunkt analysert med prinsipale komponenters metode (PCA) med ortogonal (varimax) rotasjon.

2.5.2 Beregning av konsensus

Beregning av konsensus kan gjøres på flere måter og med forskjellige effektmål, men variansmål er helt klart å foretrekke (Kenny, 1994; Shrout, 1993; Funder, 1999).

Data ble samlet inn i et fullstendig krysset design der 8-11 personer vurderte hverandre med 5PF20a. Data ble organisert i 20 matriser (20 trekk) per gruppe med gruppedeltakerne som både bedømmer og target. Diagonalene i matrisene (egenvurdering) utelates fordi hensikten er å undersøke reliabilitet for personlighetsvurdering ved første møte, og ved et senere tidspunkt når kjennskap mellom personer er større. Den totale variansen i matrisene kan deles i tre deler: den delen som skyldes target (targetvarians), den delen som skyldes bedømmer (bedømmervarians), og den delen som skyldes interaksjon og tilfeldigheter. Beregning av konsensus kan foretas ved ordinær toveis ANOVA (analysis of variance) og påfølgende beregning av intraklassekorrelasjon (Shrout og Fleiss, 1979). Dette krever i prinsippet ortogonale (ukorrelerte) faktorer. Ukorrelerte faktorer kan ikke forutsettes i disse data, både fordi diagonalene i matrisene mangler, og fordi det er grunn til å tro at det er sammenheng mellom egen personlighet og hvordan man bedømmer andre personer (Kenny, 1994; Funder, 1999). Nødvendige "mean squares" (variansestimaterne i toveis ANOVA) kan imidlertid estimeres i det generelle tilfellet med SPSS, der eventuell kovarians mellom faktorene skilles ut.

Kenny (1986) har utviklet et eget program (SOREMO) som korrigerer variansestimaterne på grunnlag av faktisk kovarians mellom faktorer. Rent statistisk benytter SOREMO den korrekte modell for beregning av konsensus, men programmet tåler ikke "missing data" og er betraktelig mer tidkrevende (Engvik, 2007, personlig kommunikasjon). Estimering av missing data er i seg selv problematisk, av disse grunner er SOREMO mindre anvendelig enn ANOVA sammen med beregning av intraklassekorrelasjon. Et underutvalg på 19 grupper har blitt analysert med begge metoder siden det er av betydelig interesse å undersøke hvor stor forskjell det er på metodene for reelle data. Resultatene (gjengitt i appendiks A) viste at forskjellene er minimale, fra en praktisk synsvinkel gir de to metodene identiske estimater av konsensus (og reliabilitet).

Jeg har derfor valgt å benytte toveis ANOVA og beregning av intraklassekorrelasjon. Innen psykometri foreligger det en lang tradisjon for beregning av reliabilitet, Shrout og Fleiss (1979) har redegjort for en rekke reliabilitetsindekser. De har følgende definisjon av intraklassekorrelasjon:

”The ICC is the correlation between one measurement (either a single rating or a mean of several ratings) on a target and another measurement obtained on that target”.
(Shrout & Fleiss, 1979, s 422).

Som diskutert i første del av oppgaven er det en teoretisk nødvendighet å anta at en bedømmelse består av flere komponenter. Med utgangspunkt i generaliserbarhetsteori (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) kan variasjon i et måledesign redegjøres for ved følgende lineære modell:

Bedømmelse = konstant + øynene som ser (bedømmereffekt) + personen som sees (targeteffekt) + interaksjon mellom bedømmer og target (interaksjonseffekt) + rest (Shrout 1993).

I variansterminologi tilsvarer dette følgende:

Total varians = Targetvariens + bedømmervariens + systematisk relasjonell varians + tilfeldig varians.

Datainnsamlingen og måledesignet legger føringer for hvilken formel for ICC som skal anvendes. I dette designet har alle personer vært både bedømmer og target, noe som fordrer toveis fremfor enveis variansanalyse. Videre kan det antas at alle bedømmere er tilfeldig tildelt i grupper fra en større populasjon av bedømmere, og de samme bedømmere benyttes, det skjer ingen utskiftning. Med disse forutsetningene skal reliabilitetsindeksen som Shrout og Fleiss (1979) har kalt for ICC 2.1, benyttes. Konsensus (og reliabilitet) er direkte relatert til hvor stor targetvariens er i forhold til totalvariens.

ICC 2.1 estimeres ved følgende formel:

$$\frac{\text{BMS-EMS}}{\text{BMS}+(k-1)\text{EMS}+k(\text{JMS-EMS})/n}$$

WMS = between targets mean square

EMS = residual (interaction+error)

JMS = between judges mean square

k= antall bedømmere

n= antall targets

Som det fremgår av denne beregnes det tre varianskomponenter, og bedømmervarians skilles ut. De blir altså kontrollert for det Cronbach (1955) kalte "elevation".

2.5.3 Anmerkning vedrørende fortolkning av intraklassekorrelasjon

Før resultatene presenteres, gjenstår et sentralt problem med denne måten å beregne reliabilitet på som krever oppmerksomhet. De teoretiske grensene for ICC er mellom 0 og 1, fordi negative varianser ikke gir noen åpenbar mening (Lahey, Downey & Saal, 1983). De virkelige grensene for ICC2.1 er en funksjon av de ANOVA-termene den inneholder. EMS inneholder både interaksjon mellom bedømmere og targets, og feil. Disse to kan ikke skilles fra hverandre, og det blir en teoretisk diskusjon om hvor stor del av EMS som skal tilskrives interaksjon. En negativ ICC kan bare oppnås dersom targetvarians er mindre enn interaksjonen (EMS). Implikasjonen her er at ICC-verdier ikke kan være signifikante dersom ikke hovedeffekten (targetvarians, BMS) er signifikant. Altså vil negative verdier for ICC2.1 ha sin bakgrunn i at variansen av interesse ikke er signifikant. Et ytterligere element er at alle effekter er estimer, og dermed utsatt for tilfeldigheter. Hvis det i realiteten ikke er noen konsensus, så vil det i halvparten av tilfellene se ut som om konsensus er negativ. Det er derfor rimelig å sette negative ICC verdier til 0.

Dette er en konservativ strategi, selv om gjennomsnitt for ICC2.1 over grupper følgelig blir høyere enn når negative varianser tas med. Ikke-signifikante targeteffekter kan oppstå av forskjellige grunner (Lahey et al., 1983). Den første grunnen er den viktigste og mest sentrale for å kunne kalle strategien konservativ. Når target effekten ikke er signifikant (noe som indikerer manglende reliabilitet i data) kan det komme av interaksjonseffekter mellom targets og bedømmere. Ved førsteinntrykk er det vanskelig å se hva disse interaksjonseffektene skulle komme av, siden det er så godt kontrollert for kjennskap. Ved vurdering av kjente er situasjonen annerledes, og fortolkningen av manglende signifikant targeteffekt blir en annen. I løpet av kurset har det utviklet seg en rekke relasjoner i gruppene, noen har lite med hverandre å gjøre, mens andre skriver oppgave sammen, etablerer studiegruppe og blir nære venner. Ved negativ ICC ved vurdering av kjente er det rimelig å tenke seg at manglende

signifikante targeteffekter kan skyldes interaksjon (det Kenny kaller "Opinion"), og faktiske korrelasjoner i matrisene maskeres av bedømmelser som har lav reliabilitet⁷.

Den andre grunnen gjenspeiler et homogent utvalg, at det rett og slett ikke eksisterer reelle forskjeller mellom targets. Konsensus forutsetter variasjon, noe som gjør korrelasjon til et uegnet mål når faktisk variasjon er liten eller fraværende. Den tredje grunnen til at target effekt ikke er signifikant, er det som knytter seg til faktisk manglende reliabilitet. Det innebærer at det ikke er noen korrelasjoner mellom bedømmelser. Som nevnt kan dette gi negative verdier for ICC2.1, men den teoretisk meningsfulle fortolkningen er fraværende konsensus, altså verdien 0.

⁷ Når det er mistanke om dette, kan "the characteristic root test of the interaction" benyttes (Lahey et al., 1983). Fortolkningen av denne blir økende kompleks ettersom antall bedømmere og targets øker, særlig når man ikke har en a priori ide om hvilke dyader som maskerer signifikante targeteffekter. Disse sammenhengene vil derfor ikke bli undersøkt her.

3. Resultater

Estimatene baserer seg både på gjennomsnitt av ICC2.1 og tilbaketransponerte korrelasjoner fra Fishers z-transformasjon, da det i litteraturen for øvrig er variasjon i hvilke estimer som benyttes. Det er i seg selv interessant og se hvor mye de beregningene skiller seg fra hverandre. Det har blitt påpekt at både z og r er unøyaktige estimer i hver sin retning (Schmidt, Hunter & Raju, 1988). Samtidig finnes det ikke noen anerkjent metode for å beregne standardfeil for tilbaketransponerte gjennomsnitt for Fishers z. Jeg har derfor brukt vanlige gjennomsnitt i figur 1 og figur 2 slik at noen aspekter ved hypotesetestingen blir visualisert. Fullstendige tabeller med gjennomsnitt, standardfeil og z-transformerte gjennomsnitt er gjengitt i appendiks B.

3.1 Faktorstruktur og begrepsvaliditet

Ved førsteinntrykk var 60 prosent varians forklart ved fem faktorer. Det optimale er at roterte faktorer forklarer omtrent like mye varians. Følgende ble funnet:

Faktor I	10,6 prosent
Faktor II	12,8 prosent
Faktor III	11,7 prosent
Faktor IV	12,7 prosent
Faktor V	12,0 prosent

Tabell 1 viser trekkenes ladninger på faktorene. Trekk som hadde mindre ladning enn .350 på faktorene er utelatt fra tabellen.

For vurdering av kjente var 71 prosent varians forklart ved fem faktorer. Den roterte faktorene forklarte følgende varians:

Faktor I	14,1 prosent
Faktor II	14,4 prosent
Faktor III	14,9 prosent
Faktor IV	13,7 prosent
Faktor V	13,7 prosent

Tabell 2 viser trekkenes ladning på faktorene. Trekk som hadde mindre ladning enn .267 er utelatt fra tabellen.

Tabell 1: Rotert Faktorladningsmatrise(a) for førsteinntrykk

Trekk	Faktor				
	I	II	III	IV	V
Aktiv	,603				
Pratsom	,803				
Dominerende	,508	-,540			
Utadvendt	,725				
Varm		,730			
Ikke kynisk		,730			
Omsorgsfull		,748			
Vennlig		,707			
Velorganisert			,691		
Strukturert			,799		
Planmessig			,725		
Disiplinert			,768		
Ubekymret				,743	
Ikke hårsår				,617	
Ikke nervøs				,798	
Ikke anspent				,758	
Original					,639
Fantasifull					,717
Kreativ					,789
Oppfinnsom					,775

a) Rotation converged in 6 iterations.

Tabell 2: Rotert faktorladningsmatrise(a) for vurdering av kjente

Trekk	Faktor				
	I	II	III	IV	V
Aktiv	,805				
Pratsom	,860				
Dominerende	,750				
Utadvendt	,769				
Varm		,832			
Ikke kynisk		,761			
Omsorgsfull		,839			
Vennlig		,810			
Velorganisert			,823		
Strukturert			,879		
Planmessig			,839		
Disiplinert			,825		
Ubekymret				,802	
Ikke hårsår				,736	
Ikke nervøs				,837	
Ikke anspendt				,826	
Original					,698
Fantasifull					,803
Kreativ					,830
Oppfinnsom					,811

a) Rotation converged in 5 iterations.

3.2 Observerbarhet og konsensus

De to kommisjonenes skalaer på observerbarhet (jamfør del 2.3) hadde en reliabilitet på .91. Konsensus og observerbarhet viste en signifikant og relativt sterk korrelasjon ved førsteinntrykk ($r=.47$, $p<.05$). For observerbarhet vurdert på skala fra 1-7 var gjennomsnittlig observerbarhet på 4.76 og standardavvik på 1.04. Minste verdi var på 2,85 (for trekket kynisk) og største verdi var på 6,79 (for trekket pratsom). Hele skalaen ble altså ikke benyttet (variasjonsbredde på 3.94). Korrelasjonene mellom konsensus og observerbarhet ved alle tidspunkt var:

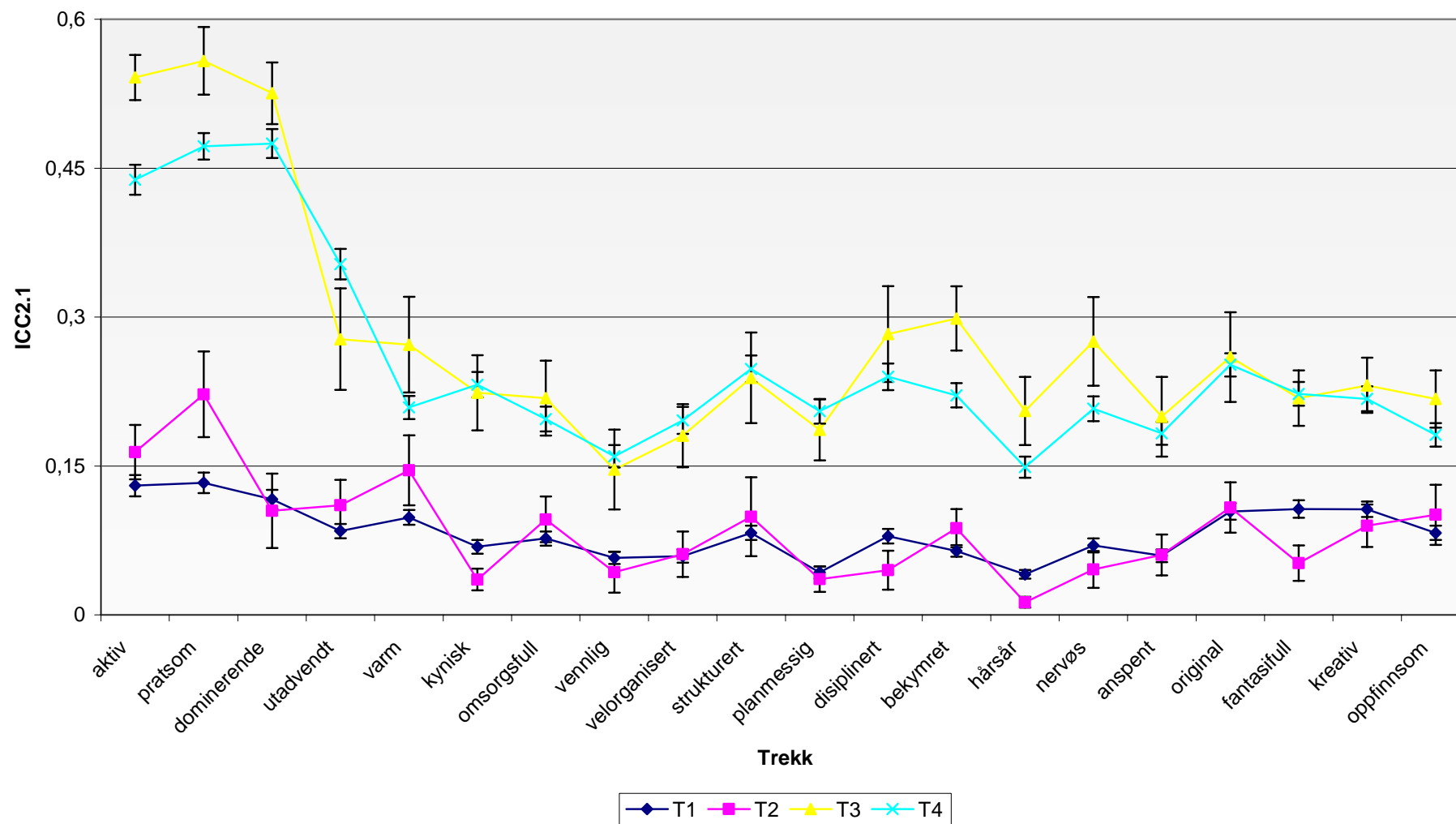
T1 (Førsteintrykk)	$r =.47$ ($p<.05$)
T2 (Ti sek presentasjon)	$r =.56$ ($p<.05$)
T3 (4 dager)	$r =.61$ ($p<.01$)
T4 (7 mnd)	$r =.69$ ($p<.01$)

Skalaen for observerbarhet er gjengitt i appendiks C.

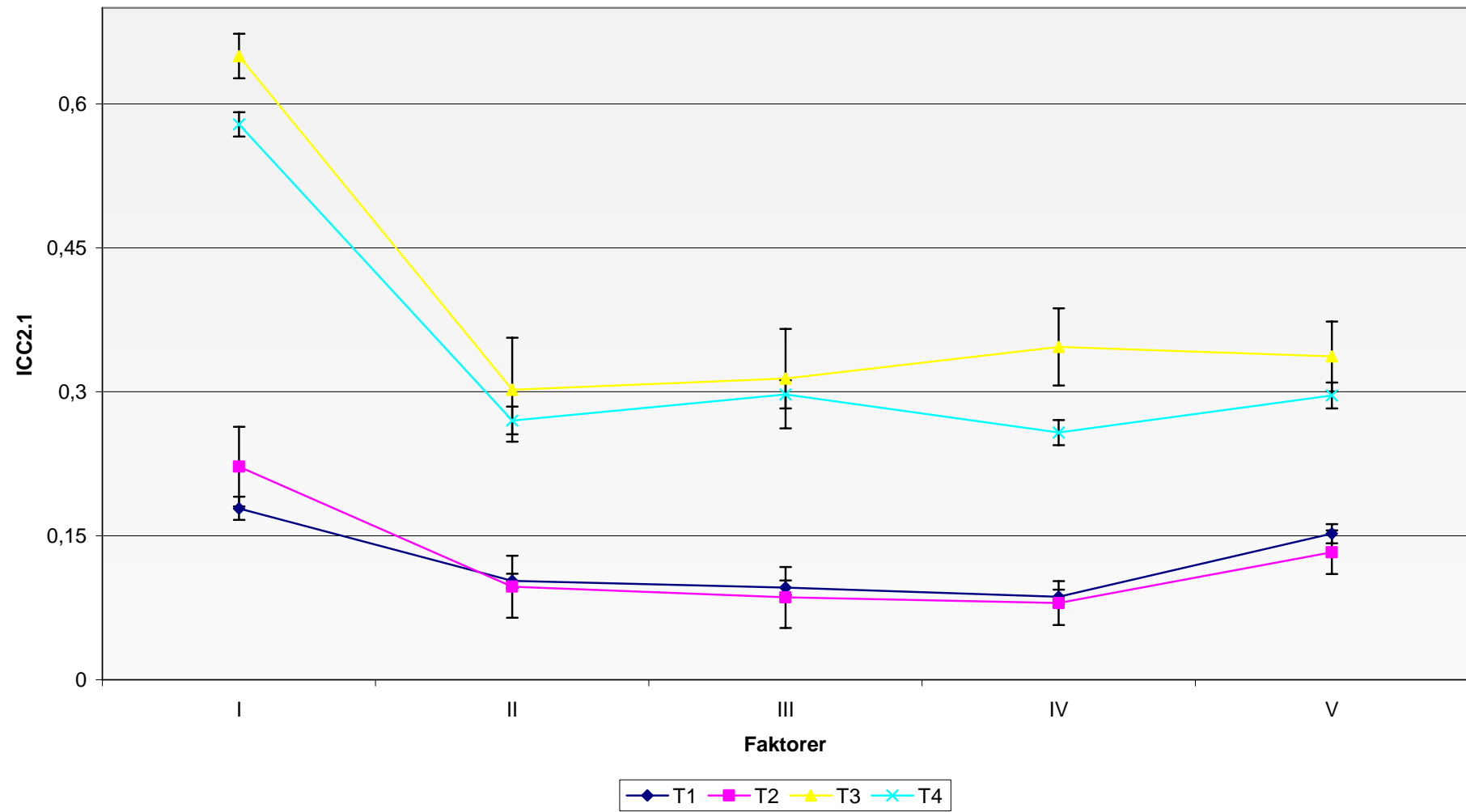
3.3 Konsensus som funksjon av kjennskap

Figur 1 viser konsensus for trekk med fire variasjoner i kjennskap: Førsteintrykk (T1), 10 sekunders presentasjon (T2), 4 dagers kjennskap (T3), og vurdering av kjente etter 7 måneder (T4). Den første horisontale referanselinjen i figuren representerer Kennys 15 prosent grense for relativ target varians (konsensus). Figur 2 tilsvarende figur 1, men trekk er byttet ut med faktorer.

Figur 1: Konsensus for trekk



Figur 2: Konsensus for faktorer



4. Diskusjon

Hovedformålet med dette studiet var å undersøke variasjon i konsensus som funksjon av observerbarhet og kjennskap. Jeg vil i det følgende gi en oppsummering av funnene, og diskutere resultatene opp mot hypotesene og relevant teori og empiri. Videre vil jeg drøfte mulig svakheter ved dette forskningsdesignet, før jeg til slutt oppsummerer, konkluderer og antyder implikasjoner for fremtidig forskning på konsensus og personpersepsjon.

4.1 Oppsummering av resultater

Jeg finner god støtte for de tre første forskningshypotesene, og noe støtte for den siste.

Faktorstrukturen var tilfredsstillende både for kjente og ukjente, med høy strukturell validitet og begrepsvaliditet. For trekk som er observerbare var det betydningsfull konsensus ved førsteinntrykk, for ekstraversjon berøres Kennys 15 prosentgrense. Videre øker konsensus som funksjon av kjennskap. For personer som kjenner hverandre er det trivielt enkelt å tilbakevise Kennys påstand om at konsensus stopper på .15, med opp mot .57 for trekk og .65 for faktorer. Det var begrenset støtte for hypotese 4, noe som fordrer en diskusjon på når i kjennsapsprosessen høyeste nivå av konsensus oppnås, om konsensus blir lavere ettersom kjennskap øker, og hva dette representerer.

4.2 Faktorstruktur er lik ved begge tidspunkt

God støtte for denne hypotesen. Den strukturell validiteten var høy ved førsteinntrykk og enda høyere for vurdering av kjente. Videre var det god begrepsvaliditet for 5PF20a, der alle adjektiver for vurdering av kjente har høy ladning på forventet faktor og liten eller fraværende ladning på de andre faktorene. Ved førsteinntrykk "taper" faktor I noe av sin varians til faktor II, på grunn av at "dominant" lader høyt på begge. Dette impliserer at disse to faktorene ikke er fullstendig ortogonale i en optimal rotert løsning, altså et uklart intersubjektivt grunnlag for hva trekket "dominant" representerer ved vurdering av ukjente. Samsvar mellom bedømmers benyttelse av trekkbeskrivelsene er en viktig side ved personlighetsbeskrivelsenes validitet. Med unntak av "dominant" for vurdering av ukjente, er individuelle forskjeller beskrevet med 5PF20a i stor grad intersubjektive faktum. Begrepsvaliditeten her var mye bedre enn tilsvarende skaler som den kan sammenlignes med (for eksempel Norman, 1963; Passini & Norman, 1966). Dette var forventet fordi trekkene ble valgt for å optimalisere

intersubjektivitet. Intersubjektive faktum kan ha svært objektive konsekvenser (jamfør Valla-saken). I et fullstendig kryss design som ble benyttet i dette studiet, kan manglende konsensus med stor sikkerhet tilskrives forskjeller i hvordan bedømmere vurderer targets. Forskjeller i konsensus over trekk og faktorer har ikke sitt utgangspunkt i uenighet om språket vi bruker for å beskrive, men i hvordan targets faktisk beskrives med språket. Selve språket (5PF20a) har stabil intersubjektiv mening.

4.3 Konsensus for ukjente er substansiell for trekk som er observerbare

Resultatene viste en signifikant og sterk sammenheng mellom konsensus ved førsteinntrykk og observerbarhet, noe som gir god støtte til hypotesen. Både for trekk og faktorer var det høyest konsensus for ekstraversjon, noe som tilsvarer funn i en rekke andre studier (Albright, Kenny, & Malloy, 1988; Borkenau & Liebler, 1992; Paunonen, 1989; 1991; Kenny et al., 1992; Watson, 1989; Funder & Dobroth, 1987; Funder & West, 1993). Hovedforskjellen mellom disse studiene var ikke de empiriske funnene, som alle går i samme retning, men forklaringene som ble benyttet. I diskusjonen frem mot hypotesen i del 1.3 ble to sentrale forklaringsmodeller berørt: a) Konsensus for ukjente representerer sosialt delte stereotypier mellom bedømmere som ikke reflekterer reelle korrelasjoner mellom fysisk fremtreden og personlighet, og b) konsensus for ukjente representerer at det finnes en viss ”kjerne av sannhet” i førsteinntrykk. I studier der den første forklaringsmodellen foretrekkes (Albright et al., 1988; Kenny et al., 1992), berøres også direkte sammenhenger mellom fysisk fremtreden og personlighet, men modellene diskuteres ikke opp mot hverandre. “The issue remains open, however, as to why there are shared stereotypes for some traits, such as sociable, but not for others” (Albright et al., 1988, s 393). Det blir nevnt at det kan være en kjerne av sannhet i førsteinntrykk, men dette blir ikke utforsket videre under argument om at designet ikke gjør det mulig å skille mellom disse to forklaringsmodellene (for eksempel Kenny et al., 1992). Det synes meg underlig å konkludere slik. Vanlig praksis i utveksling mellom teori og empiri er å knytte mest lit til de teorier som forklarer flest sammenhenger i materialet. Fra et vitenskapsteoretisk ståsted er det jo nettopp dette som driver forskningen fremover.

Er så ”kjerne av sannhet”- hypotesen i bedre stand til å forklare de empiriske funnene innen dette paradigmet? Berry og Finch Wero (1993) presenterer studier som gir spesifikk støtte til at bedømmere ansiktsuttrykk er informativt og gir valide tegn på personlighet. De finner også generell støtte til at fremtreden, holdning, stemmekvalitet og bevegelser er rike kilder til

informasjon om targets. Borkenau og Liebler (1992) foreslår at fremmede fokuserer sin begrensede informasjonsprosesseringskapasitet på de trekk som *mest nøyaktig* kan sluttes fra de fysiske attributter. I studiene ovenfor finner man nest høyest konsensus for faktor III, mens i dette studiet finner jeg det for faktor V. Hvordan skal dette kunne forklares? Albright og medarbeidere (1988) finner i ett av sine studier tendenser til det samme, og forklarer dette med egenskaper ved utvalgene. De besto av studenter ved privatuniversiteter som rimeligvis kan tenkes å verdsette intellekt høyere enn andre utvalg. Dette bemerkes i en fotnote, men det er grunn til å tro at de også her tenker seg at det finnes en sosial stereotypi som er ansvarlig for den relativt høye konsensusen rundt faktor V (Albright et al., 1988).

I tråd med det utvidede evolusjonsperspektivet som ble presentert i del 1.2.3 og 1.2.4, vil jeg foreslå en annen forklaring. I situasjoner der personer ikke kjenner hverandre, finnes det en rekke relevante sosialadaptive problemer som *kan* aktiveres i samspillet. *Når interaksjonen forsøkes holdt på et minimum, vil selve prosedyren få avgjørende betydning for hvilke sosialadaptive mekanismer som aktiveres.* Før studentene deler seg i grupper, skal de hver for seg prioritere fem farger og begrunne valgene sine. Studentene vet ikke formålet med denne prioriteringen. Selve fargeprioriteringsprosessen kan godt betegnes som en oppfordring til å være kreativ. Hvis man forstår fordelingen til grupper som et spill der sosial status og aksept utveksles, er det slik at noen må gi tapt og gå fra ”populære” farger til mindre populære farger. Disse aspektene gjør det rimelig å tenke seg et ”sosialadaptivt landskap” der jakten på sosial status og aksept befinner seg innenfor en ramme der kreativitet (Faktor V) fremstår som viktig for studentene.

Samtidig kan det godt hende at ledere som gruppe verdsetter kreativitet og originalitet høyere enn andre grupper, men det kan ikke være en sosial stereotypi om dette som er ansvarlig for den høye konsensusen. Det avgjørende i dette argumentet ble diskutert i 2.5.3: Korrelasjon forutsetter variasjon. Dersom kreativitet og originalitet betegner ledere som gruppe, innebærer det at de skulle være mer homogene på denne faktoren. Det ville gi mindre variasjon, og følgelig lav konsensus. Siden konsensus er til stede, vil en fastholdig av sosiale stereotypiers betydning måtte ta med i betraktning at ulike personer aktiverer forskjellige sosiale stereotypier. Vi ligger da tett opptil perspektivet jeg argumenterer for: observerbare individuelle forskjeller er ansvarlig for at personer også vurderes forskjellig, og at det dermed oppnås konsensus omkring hvilke egenskaper personer har.

Flere forfattere har bemerket et ytterligere problem med sosiale stereotypier (Paunonen, 1989; 1991; Berry & Finch Wero, 1993; Watson, 1989). Hvis det ikke er faktiske personlighetsforskjeller mellom personer som smiler mye og personer som smiler lite, hvorfor har det da på ett eller annet tidspunkt oppstått en sosial stereotypi som sier at de er henholdsvis ekstravert og introvert? Under innsamlingsprosedyren er det vanlig at personer ikke fullt ut klarer å forholde seg til instruksjonen om å ikke snakke sammen, et poeng Watson (1989) merket seg: "Even though extraneous talking and social interaction were actively discouraged, one experimenter reported that it proved almost impossible to keep certain subjects (presumably extraverts) quiet" (Watson, 1989, s 122).

Wiggins og Trapnells (1996) prinsipielle forrang til faktor I gir god mening til det mest stabile funnet i alle studiene jeg har diskutert her. Konsensus rundt faktor I er alltid høyest i vurdering av andre, fordi sosial status har så avgjørende betydning at mekanismene - selv under direkte instruksjoner om å ikke samhandle på vanlig måte - faktisk ikke kan "temmes". Den grunnleggende sosiale orienteringen vil alltid være aktivert i et miljø som til enhver tid tilbyr begrensede ressurser.

4.4 Konsensus er høyere for vurdering av kjente

Jeg finner god støtte for denne hypotesen. Hvis vi sammenligner T1 og T4, er det tydelig at det ikke er noen overlapp for noen trekk, konsensus er signifikant høyere for alle trekk ved de to tidspunkt. Samtidig er konsensus for alle trekk, med unntak av "hårsår", signifikant høyere enn Kennys 15 % grense. En gjennomgang av meta-analyser (Kenny, 1994; Kenny, Albright, Malloy, & Kashy, 1994) gjør det vanskelig å finne hva denne grensen egentlig baserer seg på. Det kan synes som Kenny i utgangspunktet har bestemt seg for at relativ target varians (konsensus) skal være lavere enn bedømmereffekt og interaksjonseffekt. For vurdering av kjente er det trivielt enkelt å tilbakevise hans påstand, med konsensus for faktor I på opp mot 65 prosent.

Det er interessant å se nærmere på hvordan kjennskap virker inn på konsensus. Da blir moderatoren for nøyaktighet som Funder (1999; 2001) har kalt for "good information", sentral. Det er to aspekter ved good information: informasjonens kvantitet og informasjonens kvalitet (Funder, 1999). Det er her på plass med en gjennomgang av hva tidligere forskning har vist. Ytterst få studier har fokusert på det kvalitative aspektet (Letzring, Wells & Funder,

2006), i alle fall har ikke kjennskap i særlig grad blitt differensiert i forhold til dette. Kenny og medarbeidere (1994) foretok en meta-analyse av 32 studier på konsensus. De delte studiene inn etter grad av kjennskap, og fant fem studier som de gav betegnelsen "long-term acquaintance". Konsensusnivået fra disse studiene er relativt homogent, med et gjennomsnitt på 27,5 prosent for faktorer. De finner også at nivået for faktor I er bare så vidt høyere enn ved "zero acquaintance". Altså ingen betydningsfull økning i konsensus for faktor I (Kenny et al., 1994). På treknivå er det ikke gjengitt noen resultater.

De fant videre 12 longitudinelle studier, av disse var 6 studier i klasserom, mens 6 var laboratoriestudier. Klasseromssituasjonen er mest relevant å sammenligne de foreliggende resultatene med, jeg vil gå nærmere inn på disse studiene. Et problem er at flere av disse ikke er publisert, så jeg må for det meste forholde meg til meta-analysen, og Kennys (1994) bok der ulike aspekter ved studiene er gjengitt. Jeg vil først trekke frem et poeng som ikke relaterer seg til kjennskap, men til operasjonalisering av personlighet. Tre studier benytter 5 trekk, mens en studie kun benytter to trekk som begge knytter seg til faktor I. Det er derfor god grunn til å stille spørsmålstegn ved den strukturelle validiteten i disse studiene, det er ikke gjengitt hvor mye varians som er forklart ved fem faktorer ved de ulike måletidspunktene. De to siste studiene benytter 10 trekk for fem faktorer, fortsatt halvparten av operasjonaliseringen i 5PF20a. Da det ikke her heller er gjengitt noe om den strukturelle validiteten og stabiliteten til språket som er benyttet, er det vanskelig å si i hvilken grad den relativt større økningen i konsensus i mitt studium knytter seg til dette. Siden 5PF20a er optimalisert i forhold til disse aspektene (jamfør del 4.2), vil jeg likevel påstå at man kan oppnå høyere, og ikke minst mer reliable estimer for konsensus når operasjonaliseringen av personlighet er gjennomarbeidet og testet.

I meta-analysen (Kenny et al., 1994) fant de at det samlet var en antydning til økning i konsensus i klasseromsstudiene. Faktor I og faktor II viste ikke noe økning over tid. De foreslår at den utvidede interaksjonen i klasserom kan være ansvarlig for at konsensus synes å øke noe for faktor III, IV og V (Kenny et al., 1994). Med henblikk på faktor I er det i tre av studiene en tidstrend: Konsensus er målt ved tre tidspunkt, og faktor I går opp ved det midterste tidspunktet, for så å gå ned igjen. For Faktor II er det en lignende tidstrend i 2 av studiene. Det bemerkes at det er et inkonsistent endringsmønster, men det diskuteres ikke videre i meta-analysen (Kenny et al., 1994).

I en slik diskusjon vil informasjonens *kvalitet* være sentral. Letzring og medarbeidere (2006) har spesifikt manipulert denne variabelen. De opererte med flere avhengige variabler, deriblant konsensus. De hadde som hypotese at konsensus vil være høyere i situasjoner der deltagerne interagerer under betingelser der mer personlighetsrelevant informasjon gjøres tilgjengelig (Letzring et al., 2006). Datainnsamlingen var organisert i grupper på tre og tre personer, og personlighetsvurdering ble gjort med Q-sort bestående av 100 deskriptive påstander om personlighet. Det var tre hovedvariasjoner i hvilken kvalitativ informasjon deltagerne fikk om hverandre i løpet av en 50 minutters seanse. I den første situasjonen skulle gruppen arbeide sammen om å besvare en quiz, før de vurderte hverandres personlighet. I den andre situasjonen var instruksjonen at de kunne holde på med hva de ville, mens personer i den siste situasjonen fikk beskjed om å bli så godt som mulig kjent med hverandre og finne ut så mye som mulig om hvilken type personer de tilbragte de 50 minuttene sammen med. Studiet viste at konsensus økte mellom situasjon 1 (gjennomsnittlig $r = .29$) og situasjon 2, mens konsensus hadde et stabilt nivå mellom situasjon 2 og 3 (gjennomsnittlig $r = .41$). Dette er i tråd med prediksjonen fra WAM (Kenny, 1994), som uttrykker at konsensus vil nå et maksimumsnivå (Letzring et al., 2006). I dette tilfellet var gjennomsnittets konsensus .41, nyansene på faktor- og trekknivå er ikke redegjort for.

I de fleste klasseromsstudiene har subjektene ikke interagert i mer enn 2-3 timer til sammen (Kenny et al., 1994), og operasjonaliseringen av personlighet ser ikke ut til å ha vært særlig gjennomarbeidet. I studiet til Letzring og medarbeidere (2006) var interaksjonen godt kontrollert, og de benyttet en omfattende og godt validert personlighetsoperasjonalisering. Det er derfor grunn til å feste mest lit til funnene gjort i deres studium. Konsensus øker over tid, og viste seg å ha en *gjennomsnittelig* takeffekt på 41 prosent.

Det relativt høyere konsensusnivået i mitt studium relaterer seg til bruk av gode, intersubjektive operasjonaliseringer av trekk, og at personer har hatt anledning til å interagere over mye lengre tidsrom enn i mye av forskningen ovenfor. I tillegg diskuterer jeg funnene på trekk- og faktornivå, noe som åpner opp for å se nærmere på *variasjon*, og *evolusjon*. Jeg vil nå gå over til en mer inngående diskusjon av tidstrender jeg finner resultatene.

4.5 Maksimumsnivå for konsensus

Denne hypotesen uttrykte at konsensus fortsetter å øke ettersom mer (kvantitativ og kvalitativ) informasjon om targets blir tilgjengelig. Jeg finner noe støtte for hypotesen for enkelte trekk når ulike variasjoner vurderes opp mot hverandre. I de gruppene der det var anledning til å presentere seg i ti sekunder, finner jeg signifikant høyere konsensus for trekkene ”pratsomhet” (I) og ”varme”(II), mens konsensus er signifikant lavere for trekkene ”kynisk”(III), ”disiplinert”(III) og ”hårsår”(IV), sammenlignet med T1 (ingen presentasjon). Dette er i tråd med Wiggins og Trapnells (1996) perspektiv. De signifikante forskjellene forsvinner når trekkene aggregeres til faktorer. Jeg vil i det følgende konsentrere meg om forskjellene mellom T3 (vurdering etter fire dager) og T4 (vurdering etter syv måneder).

Konsensus er høyere for enkelte trekk ved 4 dager enn ved 7 måneder. På faktornivå er konsensus signifikant høyere for faktor I og faktor IV ($p < .05$) ved vurdering etter fire dager. På trekknivå er det ”aktiv”, ”pratsom” og ”dominerende” som er ansvarlig for høyere faktor I, ”utadvendt” er signifikant lavere. ”Bekymret”, ”hårsår” og ”nervøs” utgjør det meste av forhøyningen på faktor IV. For de resterende trekk og faktorer er det ingen signifikante forskjeller, noe som tyder på at konsensus for faktor II, III og V er stabil fra 4 dager til 7 måneder. I forrige seksjon refererte jeg til noen parallelle tidstrender i klasseromsstudier, det er altså tilgjengelig evidens for at konsensus har en tilsynelatende inkonsistent variasjon over tid (Kenny et al., 1994).

Det mest robuste funn i forskningen som er presentert så langt, er at faktor I alltid kommer høyest ut. Hvordan kan det ha seg at den går ned over tid? Kenny og medarbeidere (1994) foreslår at informasjon omkring ekstraversjon prosesseres med minimal interaksjon. Ekstraversjon plasseres under Digmans (1997) overordnede faktor Beta. Hvis vi skal se dette som et dypt alvorlig (evolusjons-) spill, er Beta konkurranse, mens alpha er allianse. Altså er Beta i utgangspunktet ikke relasjonell, slik alfa er. Det at trekket ”utadvendt” er lavere ved T3 enn ved T4 støtter denne forståelsen, i og med at det – i alle fall rent språklig – har et mer relasjonelt aspekt enn de tre andre trekkene som utgjør faktor I. Dette må sees som tendenser i materialet, jeg vil trekke frem et annet element som underbygger dette, og som også kan gi forklaring til at faktor IV går ned.

Her er det igjen relevant å trekke inn hvilken type informasjon vi kan tenke oss at personer sitter inne med om personene i sin egen gruppe. Moxnes (2006) redegjør for hvordan

modulene i kurset på BI er lagt opp. I den første modulen er det sterk konkurranse mellom gruppene, som er de samme som utgjør rammen for personvurdering. De utfører en rekke oppgaver under tidspress og med begrensede ressurser, og målet er å bli den mest suksessfulle gruppen (Moxnes, 2006). Konkurranseaspektet gjenspeiles ofte i gruppenes valg av navn, eksempelvis "Perpetum possible" og "Stille storm". Mellom grupper handler det om "å bli best", mens det innad i gruppene skapes "a strong sense of social identity, group belonging and strong emotional involvement" (Moxnes, 2006, s 7). Faktor IV går under Digmans (1997) alpha. Når gruppene er under høyt emosjonelt press slik tilfellet er her, er det rimelig at individuelle forskjeller i emosjonalitet også blir tydelige. Vi kan tenke oss at det innad i gruppen blir viktig å føle seg akseptert, og dermed slippe å sitte med "skylden" dersom gruppen gjør det dårlig. Individuelle forskjeller i ekstraversjon blir tydelige på grunn av den pågående statuskampen mellom gruppene.

I de påfølgende moduler etableres andre samhandlingsformer i tråd med kursets overordnede målsetninger. I modul 4 etableres en kontekst av intimitet og nærhet ved hjelp av ikke-konkurrerende smågrupper, der personlig refleksjon står i fokus. Det oppstår sterke bånd mellom studenter i denne modulen, noe som leder meg over i en siste refleksjon rundt den aktuelle hypotesen (for utfyllende beskrivelse av studieprogrammet, se Moxnes, 2006).

Innebærer lavere konsensus ved T4 at bedømmelser blir mindre nøyaktige? Her oppstår det et problem med forskningsdesignet og analysene jeg har benyttet meg av: gruppene er ikke lenger organisert i en "round-robin fashion" (Funder, 1999, s 97): personer kjenner hverandre fra forskjellige sammenhenger og i varierende grad. Fire dager er neppe nok til at tette dyader har utviklet seg, slik vi kan anta ved syv måneder. Ettersom kjennskap øker vil konsensus i mindre grad reflektere nøyaktige bedømmelser, fordi kjennskap varierer mer innad i grupper. Noen personer står rimeligvis i bedre posisjon enn andre til å vurdere en person nøyaktig, fordi de har opplevd personen i lengre tidsrom (kvantitet), og i forskjelligartede kontekster der ulike sosialadaptive problemer har vært relevante (kvalitet). Det er rimelig at slike interaksjonseffekter maskeres av mindre nøyaktige bedømmelser gjort av personer som har mindre kjennskap (jamfør del 2.5.3). Dette har riktignok ikke blitt undersøkt i dette studiet, og må bekreftes av annen forskning. Med disse metodesvakheter er det vanskelig å si hva det absolutt høyeste nivået av konsensus kan være, men det er hevet over enhver tvil at konsensus kan nå høye nivåer.

Et hovedpoeng, ut ifra evolusjonsperspektivet presentert her, er at mennesket har godt utviklede mekanismer som gjør oss i stand til å avdekke individuelle forskjeller hos andre. Konsensus varierer etter hvilke sosialadaptive mekanismer som er aktivert, og *når personlighetsmålingen ligger tett opp til situasjoner der spesifikke sosialadaptive problemer har vært aktuelle, ser det også ut til at konsensus ligger på sitt høyeste*. Variasjonene i dette studiet er ikke mange nok til å si hvor tidlig i kjennskapsprosessen dette skjer, studiet til Letzig og medarbeidere (2006) tyder på at det er mer avhengig av kvalitet enn kvantitet.

4.6 Styrker og svakheter ved forskningsdesignet

Som diskutert i forrige seksjon er det et problem å bruke dette designet når personer kjenner hverandre godt. Personer som er nære venner må antas å kjenne hverandre bedre enn personer som er bekjente, men med et round-robin design kan man ikke undersøke dette. I de teoretiske begrunnelsene for designet (WAM, Kenny, 1994; PERSON, Kenny, 2004) gjøres det ingen distinksjon mellom hvem som persiperer target. "The person perceiver is not seen as an active processor of information" (Kenny, 1994, s 211).

En annen og mer opplagt svakhet ved designet, er at det skal gjøres svært mange vurderinger av hver bedømmer. Dette vil rimeligvis trekke ned reliabiliteten. Slik sett har den menneskelige bedømmer en innebygd variasjon i reliabilitet, og andre metoder kan være i stand til å skape betingelser som tar bedre hensyn til dette faktum. Her har hensikten vært å undersøke konsensus, noe som dette designet er skreddersydd for. Som Funder uttrykker: "The appropriate design and analysis depends on the research questions beeing asked" (Funder, 1999, s 99). I det foreliggende materialet finnes også egenvurderinger med to ulike operasjonaliseringer for personlighet. Et mye brukt kriterium for nøyaktighet, er å se i hvilken grad andres vurdering sammenfaller med egenvurdering (ofte betegnet "self-other agreement", se for eksempel Funder & West, 1993). Med utgangspunkt i denne designsvakheten har både egenvurdering og vurdering av andre, fordi de er innsamlet på samme tidspunkt, lavere reliabilitet. Samtidig som dette er en svakhet, kan det i dette materialet knyttes ulike forbindelser mellom egenvurdering og andres vurdering for en stor gruppe mennesker.

Jeg vil fremheve en siste begrensning som omhandler generalisering. Utvalget består av ledere, altså en selektert gruppe som per definisjon befinner seg i de øverste deler av

samfunnet. Nå var det riktignok tilstrekkelig variasjon innad i gruppen til at konsensus oppnår imponerende høyt nivå, især for faktor I og faktor V. Det kan likevel hende at dette nettopp skyldes at sosial status er viktigere for personer som velger å bli ledere. Dette kan se ut som en sosial stereotypi, men mitt poeng er diametralt motsatt: Fordi sosial status kan tenkes å være relativt viktigere for lederpersonligheter, er de også, relativt til andre grupperinger, flinkere og mer nøyaktige i å fange opp individuelle forskjeller. For å utforske dette videre vil det være interessant etablere et lignende design for eksempel på kurs for selgere eller etterutdannelse for psykologer.

4.7 Konkluderende betraktninger

Funders modell (1999; 2001) har her vist seg å fungere godt som et overordnet rammeverk. I dette studiet har hovedfokus vært på to av hans moderatorer for nøyaktighet: "good trait" og "good information". Med bruk av en operasjonalisering av personlighet (5PF20a) som gir stabil intersubjektiv mening, har jeg ytterligere bekreftet at noen trekk *er* gode, slik all forskningen viser for ekstraversjon. Videre har jeg vist at noen trekk *blir* gode, sannsynligvis som en funksjon av at spesifikke sosialadaptive mekaniser aktiveres. Konsensus rundt faktor IV er høy når mekanismer for å oppnå sosial aksept aktiveres. Jeg har med dette studiet sannsynliggjort at kvalitetsaspektet ved "good information" også handler om individuell tilpasning til et sosialt system. Konsensus omkring faktorer og trekk avhenger av de spesifikke sosialadaptive mekanismene som aktiveres og som synliggjør individuelle forskjeller i personlighet.

"The good target" har en nær sammenheng med gode trekk (Funder, 1999), det forblir likevel et åpent spørsmål om personene i dette studiet er vanskelige eller lette å vurdere. Dette aspektet ved Funders modell har ikke fått noen spesifikk behandling i det foreliggende studiet. Det samme gjelder for moderatoren "good judge", selv om det på gruppenivå, representert ved konsensus, oppnås høye nivåer. Fra et evolusjonsperspektiv er denne moderatoren til en viss grad innebygd. Det kan tenkes at personer, av mangfoldige årsaker, lykkes i sin individuelle jakt på status og aksept. Fra perspektivet her ville det være interessant å se videre på avhengig variabel som går knytter seg til hvem som blir leder blant lederne? Gjennom kurset på BI kartlegges en rekke slike variabler som kan undersøkes videre (se Moxnes, 2006).

Det kan se ut som at Kenny (1994) har rett i at det finnes et tidspunkt der kjennskap ikke leder til høyere konsensus. Dette nivået er betraktelig høyere enn hva han sier, og lengde på kjennskap (kvantitet) synes å være underordnet kvaliteten på kjennskapet. Selve konklusjonen om at kjennskap ikke synes å være så viktig innen personpersepsjon, er direkte misvisende. Det kan se ut som at konsensus relativt raskt når et høyt nivå, slik Kenny (2004) antyder. Dette er å forvente hvis utfallet av personpersepsjon er evolusjonsmessig viktig: De sosialadaptive mekanismene er raske, effektive og mer eller mindre automatisert. En tilstrekkelig nøyaktig oppfattelse av hvordan personer rundt oss er, er av avgjørende betydning.

5. Litteraturliste

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of personality and social psychology*, 55(3), 387-395.
- Altman, I., & Taylor, D. (1987). Communication in interpersonal relationships: Social penetration theory. I M. E. Roloff & G.R. Miller (ed.), *Interpersonal processes: new directions in communication research* (s. 257-277). Newbury Park, CA: Sage.
- Berry, D. S., & Finch Wero, J. L. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of personality*, 6(4), 497-520.
- Bhaskar, R. (1989). *Reclaiming reality: A critical introduction to contemporary philosophy*. London: Verso.
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of personality and social psychology*, 62(4), 645-657.
- Buss, D. M. (1996). Social adaptation and five major factors of personality. I J. S. Wiggins (ed.), *The five-factor model of personality: Theoretical perspectives* (s. 180-207). New York: Guilford Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological bulletin*, 56, 81-105.
- Cook, T. D., & Campbell, D. T. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141-180.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding others" and "assumed similarity." *Psychological Bulletin*, 52, 177-193.
- Cronbach, L. J., Gleser, G.C., Nanda, H., & Rajaratnam, N (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: Wiley.
- Dancy, J. (1995). *An introduction to contemporary epistemology*. Oxford, UK: Blackwell Publishers.
- Darwin, C. (1859). *On the origin of the species by means of natural selection, or Preservation of favoured races in the struggle for life*. London: Murray.
- Digman, J. M. (1997). Higher-order factors of the big five. *Journal of personality and social psychology*, 73(6), 1246-1256.
- Engvik, H. (1993). "Big five" på norsk. *Tidsskrift for norsk psykologforening*, 30, 884-896.

- Engvik, H., & Føllesdal, H. (2005). The big five inventory på norsk. *Tidsskrift for norsk psykologforening*, 42, 128-129.
- Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, 98, 513-537.
- Funder, D. C. (1996). What do you think of me? *Psychological inquiry*, 7 (3), 275-278.
- Funder, D. C. (1999). *Personality judgement: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Funder, D. C. (2001) Accuracy in personality judgement. Research concerning an obvious question. I B. W. Roberts, & R. Hogan (ed.). *Personality and psychology in the workplace. Decade of behavior (s121-140)*. Washington, DC: American Psychological Association.
- Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of research in personality*, 40, 21-34.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of personality and social psychology*, 52, 409-418.
- Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgement. *Journal of personality*, 61, 457-476.
- Galton, F. (1884). Measurement of character. *Forthnightly review*, 36, 179-185.
- Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. I L. Wheeler (ed.), *Review of personality and social psychology*, (s 141-165). (2. utg.). Beverly Hills: Sage.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.
- Hogan, R. (1996). A socioanalytic perspective on the five-factor model. I J. S. Wiggins (ed.), *The five-factor model of personality: Theoretical perspectives* (s. 163-179). New York: Guilford Press.
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement and theoretical perspectives. I L. A. Pervin & O. P. John (red.), *Handbook of personality: Theory and research* (s. 102 – 138). (2. utg.). New York: Guilford Press.
- Kammrath, L. K., Mendoza-Denton, R., & Mischel, W. (2005). Incorporating if...then...personality signatures in person perception: Beyond the person-situation dichotomy. *Journal of personality and social psychology*, 88(4), 605-618.
- Kenny, D. A. (1986). SOREMO: A FORTRAN computer program for round robin data structures. Upublisert manuskript, University of Connecticut.

- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and social psychology review*, 8(3), 265-280.
- Kenny, D.A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the big five. *Psychological bulletin*, 116(2), 245-258.
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L.C. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of personality and social psychology*, 62(1), 88-97.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting for controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23-34.
- Kruglanski, A. W. (1989). The psychology of being “right”: the problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395-409.
- Lahey, M. A., Downey, R. G., & Saal, F. E. (1983). Intraclass correlations: There’s more than meets the eye. *Psychological bulletin*, 93(3), 586-595.
- Letzring, T. D., Wells, S. W., & Funder, D. C. (2006). Information quantity and quality affect realistic accuracy of personality judgement. *Journal of personality and social psychology*, 91(1), 111-123.
- Magnus, G. (2007, 10. mars). Utvalget: Valla brøt loven – Fougner-utvalget bygget på Yssens forklaring – Vallas fremstilling nærmest tilsidesatt. *Aftenposten Morgen*, s. 2.
- McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. I L. A. Pervin & O. P. John (ed.). *Handbook of personality. Theory and research* (s. 139-153). (2. utg.). New York: Guilford Press.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualising situations, dispositions, dynamics, and invariance in personality structure. *Psychological review*, 102, 246-268.
- Moxnes, P. (2006). Learning group psychology, leadership and anxiety coping by use of fairy tales as educational guide. *Nordic psychology*, 58(1), 5-21.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of abnormal and social psychology*, 66(6), 574-583.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of personality and social psychology*, 4(1), 44-49.

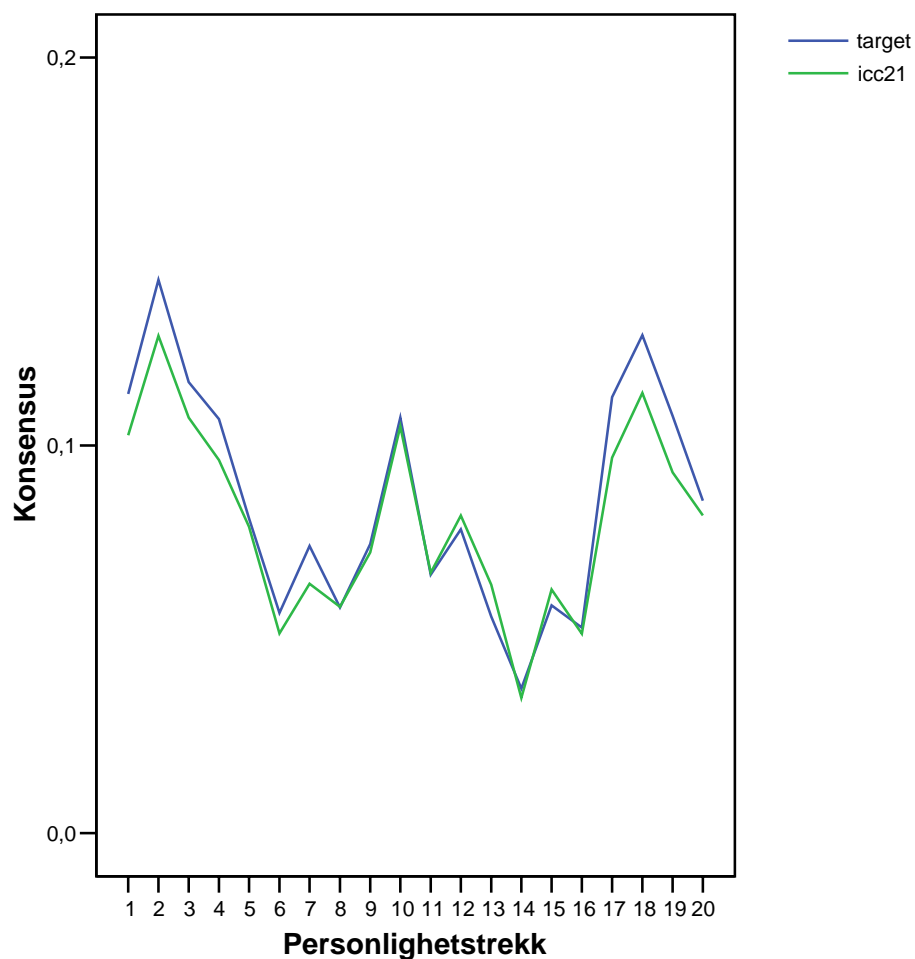
- Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of personality and social psychology*, 56(5), 823-833.
- Paunonen, S. V. (1991). On the accuracy of ratings of personality by strangers. *Journal of personality and social psychology*, 61(3), 471-477.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of educational psychology*, 74(2), 166-169.
- Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. I J. S. Wiggins (ed.), *The five-factor model of personality: Theoretical perspectives* (s. 21-50). New York: Guilford Press.
- Sampspill og ledelse: Anvendt organisasjonsledelse*. (udatert). [Internett] Tilgjengelig fra BI's webside: http://www.bi.no/templates/kurs____14334.aspx [lest 20. mars 2007]
- Saugstad, P. (1998). *Psykologiens historie – en innføring i moderne psykologi*. Oslo: Ad Notam Gyldendal.
- Schmidt, F. L., Hunter, J. E., & Raju, N. S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and Fisher's z transformation. *Journal of Applied Psychology*, 73, 665-672.
- Shrout, P. E. (1993). Analyzing consensus in personality judgments: A variance components approach. *Journal of personality*, 61(4), 769-788.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing interrater reliability. *Psychological Bulletin*, 2, 420-428.
- Watson, D. (1989). Strangers ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of personality and social psychology*, 57(1), 120-128.
- Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. I J. S. Wiggins (ed.), *The five-factor model of personality: Theoretical perspectives* (s. 88-162). New York: Guilford Press.
- Yssen, I. (2007, 10. Januar). Full strid om Vallas lederstil. Arbeidsliv. [Internett]. *Dagbladet*. Tilgjengelig fra: Atekst [lest 17. januar 2007]

6. Appendiks

6.1 Appendiks A

Figur A viser konsensus for trekk med SOREMO og ICC2.1. Gjennomsnittskonsensus (for alle trekk i 19 grupper) var .08575 for SOREMO og .08060 for ICC 2.1. Korrelasjonen mellom de to seriene var på .95. For faktorer som er sammensatt av flere trekk skiller SOREMO mellom felles og spesifikk konsensus. Konsensus for felles varians i SOREMO vil derfor gi et systematisk lavere resultat enn ICC 2.1 for sumskårer (henholdsvis .07416 og .1142 i disse data). Denne forskjellen beror imidlertid ikke på kovariansen mellom target og bedømmer og berører ikke problemstillingen vedrørende valg av analyse. De relevante forskjellene mellom metodene er minimale.

Figur A:



Konsensus for 20 trekk med SOREMO (target) og ICC 2.1

6.2 Appendiks B

Tabell B1

Trekk	T1			T2			T3			T4		
	<i>r</i>	<i>z</i>	<i>E</i>	<i>r</i>	<i>z</i>	<i>e</i>	<i>r</i>	<i>z</i>	<i>e</i>	<i>r</i>	<i>z</i>	<i>e</i>
Aktiv	0,13	0,13	0,0108	0,16	0,17	0,0274	0,54	0,54	0,0228	0,44	0,45	0,0150
Pratsom	0,13	0,14	0,0104	0,22	0,23	0,0431	0,56	0,57	0,0341	0,47	0,49	0,0133
dominerende	0,12	0,12	0,0096	0,10	0,11	0,0375	0,53	0,53	0,0311	0,47	0,49	0,0146
Utadvendt	0,08	0,08	0,0072	0,11	0,11	0,0256	0,28	0,29	0,0512	0,35	0,36	0,0152
Varm	0,10	0,10	0,0074	0,15	0,15	0,0352	0,27	0,28	0,0482	0,21	0,21	0,0117
Kynisk	0,07	0,07	0,0068	0,04	0,04	0,0110	0,22	0,23	0,0379	0,23	0,24	0,0128
omsorgsfull	0,08	0,08	0,0072	0,10	0,10	0,0230	0,22	0,22	0,0378	0,20	0,20	0,0124
Vennlig	0,06	0,06	0,0060	0,04	0,04	0,0206	0,15	0,15	0,0402	0,16	0,16	0,0114
velorganisert	0,06	0,06	0,0067	0,06	0,06	0,0229	0,18	0,18	0,0318	0,20	0,20	0,0137
strukturert	0,08	0,08	0,0072	0,10	0,10	0,0398	0,24	0,24	0,0457	0,25	0,25	0,0134
planmessig	0,04	0,04	0,0062	0,04	0,04	0,0129	0,19	0,19	0,0307	0,21	0,21	0,0126
Disiplinert	0,08	0,08	0,0074	0,05	0,05	0,0196	0,28	0,29	0,0483	0,24	0,25	0,0135
Bekymret	0,06	0,06	0,0059	0,09	0,09	0,0193	0,30	0,30	0,0324	0,22	0,23	0,0122
Hårsår	0,04	0,04	0,0044	0,01	0,01	0,0054	0,21	0,21	0,0344	0,15	0,15	0,0107
Nervøs	0,07	0,07	0,0070	0,05	0,05	0,0186	0,28	0,28	0,0447	0,21	0,21	0,0124
Anspent	0,06	0,06	0,0064	0,06	0,06	0,0206	0,20	0,20	0,0401	0,18	0,19	0,0116
Original	0,10	0,11	0,0084	0,11	0,11	0,0254	0,26	0,26	0,0452	0,25	0,26	0,0117
Fantasisfull	0,11	0,11	0,0088	0,05	0,05	0,0178	0,22	0,22	0,0280	0,22	0,23	0,0120
Kreativ	0,11	0,11	0,0077	0,09	0,09	0,0214	0,23	0,23	0,0278	0,22	0,22	0,0124
oppfinnsom	0,08	0,08	0,0072	0,10	0,10	0,0303	0,22	0,22	0,0288	0,18	0,19	0,0119

Tabell B2

Faktorer	T1			T2			T3			T4		
	<i>r</i>	<i>Z</i>	<i>e</i>	<i>r</i>	<i>z</i>	<i>e</i>	<i>r</i>	<i>z</i>	<i>E</i>	<i>r</i>	<i>z</i>	<i>e</i>
Ekstraversjon	0,18	0,18	0,0122	0,22	0,23	0,0415	0,65	0,66	0,0233	0,58	0,60	0,0127
Varme	0,10	0,10	0,0076	0,10	0,10	0,0323	0,30	0,31	0,0542	0,27	0,28	0,0144
Samvittighetsfullhet	0,10	0,10	0,0077	0,09	0,09	0,0318	0,31	0,32	0,0517	0,30	0,31	0,0147
Emosjonell Stabilitet	0,09	0,09	0,0077	0,08	0,08	0,0228	0,35	0,35	0,0402	0,26	0,26	0,0132
Intellekt	0,15	0,15	0,0099	0,13	0,13	0,0227	0,34	0,34	0,0365	0,30	0,31	0,0134

Forklaring til tabell B1 og B2

Tabell B1 er for trekk, og tabell B2 for faktorer.

T1 = førsteinntrykk N = 107

T2 = 10 sekunders presentasjon N = 9

T3 = vurdering etter fire dager N = 10

T4 = vurdering etter syv måneder N = 124

Differanse (T1,T2,T3) - T4 = -2. To grupper ble ekskludert på grunn av for mye ”missing data”.

r er gjennomsnitt av ICC2.1 for trekk for N grupper.

z er tilbaketransponerte gjennomsnitt fra Fishers z-transformasjon for N grupper.

e er standardfeil for r.

6.3 Appendiks C

Tabell C:

Trekk	Faktor	Verdi
pratsom	I	6,79
aktiv	I	6,59
vennlig	II	5,96
utadvendt	I	5,94
dominerende	I (II-)	5,73
nervøs	IV	5,25
velorganisert	III	5,21
omsorgsfull	II	4,74
anspent	IV	4,66
strukturert	III	4,6
original	V	4,54
disiplinert	III	4,53
kreativ	V	4,38
planmessig	III	4,3
varm	II	4,09
fantasifull	V	3,97
bekymret	IV	3,93
oppfinnsom	V	3,63
hårsår	IV	3,42
kynisk	II	2,85

Kommentar til tabell C:

Tabellen viser to kommisjoners vurderinger av hvilke trekk i 5PF20a som er mest observerbare. Kommisjonenes gjennomsnittsvurderinger korrelerte med .91. Vurderingene ble gjort på 7-punkts skala. Den midterste kolonnen angir hvilken faktor trekket lader på.